Algebra lineal

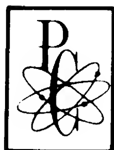
STEPHEN H. FRIEDBERG

ARNOLD J. INSEL

LAWRENCE E. SPENCE

Illinois State University

PRIMERA EDICION
MEXICO, 1982



PUBLICACIONES CULTURAL, S.A.

Contenido

Prólogo

El lenguaje y los conceptos de la teoría de matrices y, más generalmente del Algebra Lineal han llegado también a aplicarse en las ciencias naturales y en las ciencias sociales. Pero eso no priva que el Algebra Lineal continúe teniendo su importancia extraordinaria en el tratamiento moderno de la geometría y el análisis.

El propósito esencial de este libro es presentar cuidadosamente los principales temas del álgebra lineal e ilustrar la utilidad de la materia a través de una amplia variedad de aplicaciones. Aunque para el uso formal de este libro se supone que los alumnos han debido llevar un curso previo de cálculo, el contenido de los capítulos 6 y 7 no requieren más aparato matemático que el contenido en los estudios de enseñanza media superior en los cuales puede haber habido o no una iniciación al álgebra lineal.

El libro está concebido de manera tal que permite ser utilizado en cursos de diferente duración. El material esencial del álgebra lineal (espacios vectoriales, transformaciones lineales y matrices, sistemas de ecuaciones lineales, determinantes y diagonalización), se encuentra en los capítulos 1 al 5; los otros capítulos, que tratan las formas canónicas y espacios con producto interior, son completamente independientes y que se pueden estudiar en cualquier orden. Además, a lo largo del libro se encuentran diversas aplicaciones para áreas tales como ecuaciones diferenciales, economía, geometría y física. Estas aplicaciones, claro está, no son imprescindibles para el desarrollo matemático y pueden muy bien eliminarse a criterio del profesor.

Hemos procurado que resultara posible abarcar la mayoría de los temas importantes de álgebra lineal en un curso semestral. Esta meta nos permitió desarrollar los temas más importantes con menos preliminares innecesarias, que en los textos tradicionales. Nuestro tratamiento de la forma canónica de Jordan, por ejemplo, no requiere de la teoría de polinomios. La economía lograda en extensión permite desarrollar la mayor parte del libro (si se omiten muchas de las partes optativas y el análisis detallados de los determinantes), en un curso semestral de

4 horas semanales para aquellos estudiantes que hayan tenido conocimientos previos de álgebra lineal.

El capítulo 1 del libro presenta la teoría básica de espacios vectoriales de dimensiones finitas, subespacios, combinaciones lineales, dependencia e independencia lineal, bases y dimensión. El capítulo termina con una sección optativa en la cual se prueba la existencia de una base en los espacios vectoriales de dimensiones infinitas.

En el capítulo 2 se desarrollan las transformaciones lineales y sus relaciones con las matrices; ahí se discute el espacio vacío y el límite de una transformación lineal, representaciones matriciales de una transformación, isomorfismos y cambios de coordenadas. El capítulo se termina con las secciones opcionales sobre espacios duales y ecuaciones lineales diferenciales homogéneas.

En el capítulo 3 se encuentran las aplicaciones de la teoría de espacios vectoriales y transformaciones lineales a los sistemas de ecuaciones lineales. Este importante tema lo hemos pospuesto intencionadamente para que se pueda presentar como consecuencia del material anterior. Este enfoque da pie al tema familiar de los sistemas lineales para aclarar la teoría abstracta, y permite evitar confusos cálculos de matrices en los capítulos 1 y 2. En esos capítulos habrá ejemplos ocasionales donde tendremos la oportunidad de solucionar sistemas de ecuaciones lineales (naturalmente estos ejemplos no forman parte del desarrollo teórico). En la sección 1.4 se hallan las bases necesarias para ello.

Los determinantes, tema del capítulo 4, tienen ahora mucho menos importancia que hace algún tiempo, para un curso abreviado es preferible tratarlos ligeramente, puesto que consideramos necesario dedicar más tiempo a los temas que se desarrollan del capítulo 5 al 7. De ahí que hayamos presentado dos alternativas en el capítulo 4: un desarrollo completo de la teoría (secciones 4.1 a 4.4) y un resumen de los puntos importantes, indispensables para el resto de los capítulos (sección 4.5).

En el capítulo 5 se desarrollan eigenvalores, eigenvectores y diagonalización. Una de sus aplicaciones más importantes se encuentra en el cálculo de límite de matrices. Se ha incluido, sin embargo, una sección opcional sobre límite de matrices y cadenas de Markov, aunque la general mayoría de algunos de sus resultados requiera un conocimiento de las formas canónicas de Jordan. Las secciones 5.4, 5.5 y 5.6 contienen información sobre subespacios invariantes, el teorema de Cayley-Hamilton y del polinomio mínimo, respectivamente.

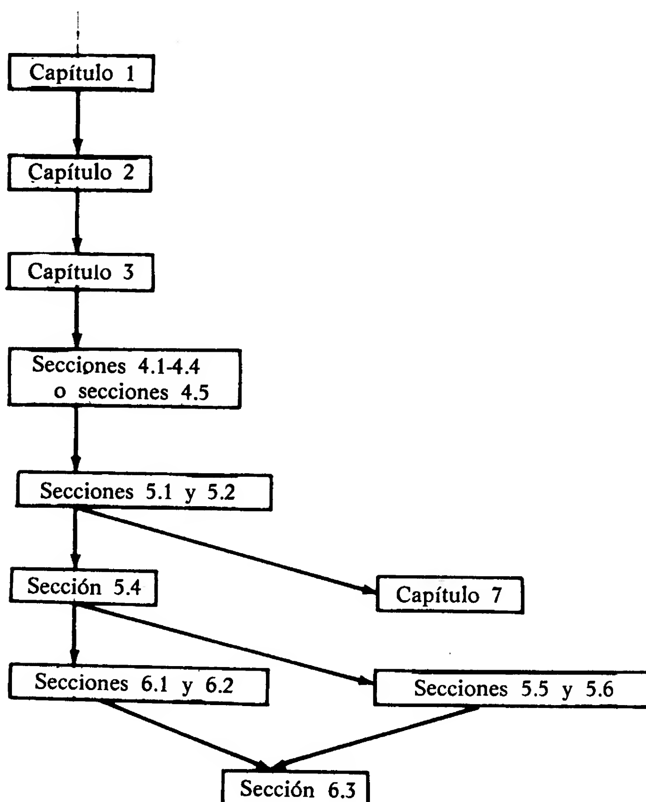
Las formas canónicas se tratan en el capítulo 6, secciones 6.1 y 6.2 desarrollan la forma Jordan y la sección 6.3 presenta la forma racional.

Los espacios con producto interior son el tema del capítulo 7. La teoría matemática básica (productos interiores y el proceso de ortogonalización de Gram-Schmidt; las transformaciones del adjunto: normal, autoadjunto, ortogonal y operadores unitarios; proyecciones ortogonales y el

teorema espectral) se desarrollan en las secciones 7.1, 7.2, 7.3, 7.5, 7.7 y en la 7.9. En las secciones 7.4, 7.6, 7.8 y 7.10 se encuentran varias aplicaciones de la estructura del producto de interior. El capítulo termina con un análisis de las formas cuadráticas y bilineales (sección 7.11).

En el texto se encuentran también cinco apéndices. En los primeros cuatro se analizan respectivamente, conjuntos, funciones, campos y números complejos con el fin de repasar las ideas básicas que se desarrollan a través del libro. En apéndice E sobre polinomios se utiliza primordialmente en los capítulos 5 y 6, en especial en la sección 6.3. Se ha preferido que esos apéndices no se analicen en forma independiente sino hacer referencia a ellos según se requiera.

El siguiente diagrama muestra la dependencia entre los capítulos del libro.



Ahora unas palabras finales, que creemos necesarias respecto a nuestra notación. Las secciones indicadas con un asterisco (*) son opcionales y pueden omitirse si así lo considera el profesor. Todo ejercicio indicado

por el símbolo (†) no es opcional; lo usamos para identificar un ejercicio que será citado posteriormente en el texto.

Agradecemos a Douglas E. Cameron *(University of Akron)*, Edward C. Ingraham de *(Michigan State University)*, David E. Kullman *(Miami University)*, Carl D. Meyer, Jr. *(North Carolina State University)* y Jean E. Rubin *(Purdue University)* por haber revisado el manuscrito completo del texto, así como también a nuestros colegas y estudiantes por las sugerencias y estímulos recibidos durante el periodo en el que se estaba desarrollando el manuscrito de esta obra. También hacemos mención especial a Miss Jana Gehrke y a Marilyn Parmantie por su ayuda en el trabajo de mecanografía, así como a Harry Gaines, Ian List y al equipo de Prentice-Hall por su colaboración durante los procesos de producción.

Normal, Illinois

STEPHEN H. FRIEDBERG
ARNOLD J. INSEL
LAWRENCE E. SPENCE

Capítulo 1

Espacios vectoriales

1.1 INTRODUCCION

Muchas nociones físicas comunes, tales como las fuerzas, velocidades * y aceleraciones, involucran una magnitud (el valor de la fuerza, velocidad o aceleración) y una dirección. Cualquier entidad que involucre magnitud y dirección se llama vector. Los vectores se representan por flechas en las que la longitud de ellas define la magnitud del vector, y la dirección de la flecha representa la dirección del vector. En la mayor parte de las situaciones físicas que involucran vectores, únicamente la magnitud y dirección del vector son significativas; consecuentemente, consideraremos a los vectores con la misma magnitud y dirección como iguales, independientemente de sus posiciones relativas.

En esta sección se discutirá la geometría de los vectores, geometría que se deriva de los experimentos físicos que dan fe de la forma de interacción entre dos vectores.

Muchas situaciones comunes sugieren que cuando dos vectores actúan simultáneamente en un punto, la magnitud del vector resultante (el vector obtenido sumando los dos vectores originales) no es necesariamente igual a la suma de las magnitudes de los dos vectores. Por ejemplo, un nadador que nada contra la corriente con una velocidad promedio de 3.2 km/h, siendo la velocidad de la corriente de 1.6 km/h, no avanzará con una velocidad promedio de 4.8 km/h. En este caso los movimientos del nadador y el de la corriente son contrarios y, por tanto, la velocidad promedio del nadador es únicamente de 1.6 km/h. Si, por el contrario, el nadador avanzara aguas abajo (a favor de la corriente), entonces su avance promedio sí sería de 4.8 km/h.

Los experimentos muestran que los vectores se suman de acuerdo con la siguiente ley del paralelogramo. (Véase la fig. 1.1.)

---

* La palabra "velocidad" está siendo utilizada con su connotación científica, como una entidad que tiene magnitud y dirección. La magnitud de una velocidad (independientemente de la dirección del movimiento) se llama *rapidez*.

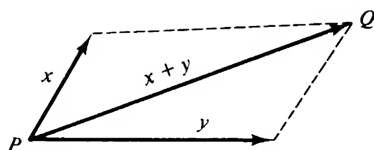


figura 1.1

Ley del Paralelogramo para la Suma de Vectores. *La suma de dos vectores* x *y* y *que actúan sobre un mismo punto* P *es el vector que, en el paralelogramo que tiene a* x *y* y *por lados adyacentes, se representa por la diagonal que parte de* P.

Como los lados opuestos de un paralelogramo son paralelos y de igual longitud, el extremo Q de la flecha que representa a $x + y$ también se puede obtener permitiendo que x actúe sobre P y luego permitiendo que y actúe sobre el extremo de x; o, de la misma manera, puede ser obtenido permitiendo que primero actúe y sobre P y posteriormente que x actúe sobre el extremo de y. De este modo, dos vectores x y y que actúan sobre un punto P pueden ser sumados "cola con cabeza"; esto es, se puede aplicar cualquiera de los vectores x o y en P y un vector que tenga la misma magnitud y dirección que el vector restante puede ser aplicado entonces en el extremo del primero —el extremo de este segundo vector es el extremo de $x + y$.

La suma de vectores puede ser descrita algebraicamente mediante el uso de geometría analítica. En el plano que contiene a x y a y, introdúzcase un sistema de coordenadas con P por origen y sea (a_1, a_2) el extremo de x y (b_1, b_2) el de y. Entonces, tal como lo muestra la figura 1.2, las coordenadas de Q, extremo de $x + y$, son $(a_1 + b_1, a_2 + b_2)$. De aquí en adelante, cuando se haga referencia a las coordenadas del extremo de un vector, se considerará que el vector parte del origen. Mas aún, como un vector que principia en el origen queda completamente determinado por

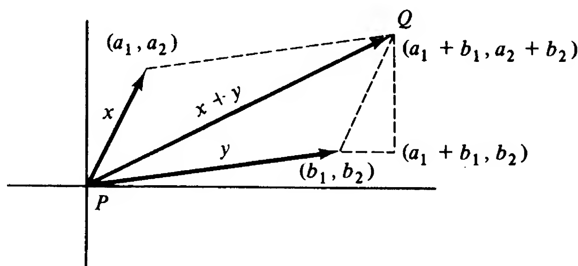


figura 1.2

las coordenadas de su punto extremo, nos referiremos algunas veces *al punto* x en vez de *al extremo del vector* x cuando x sea un vector que parte del origen.

Además de la operación de suma de vectores existe otra operación natural que se puede realizar con los vectores —la longitud de un vector puede ser amplificada o reducida sin cambiar la dirección del vector. Esta operación, llamada multiplicación por un escalar, consiste en multiplicar un vector por un número real. Si el vector x está representado por una flecha, se tiene que para cualquier número real $t \geq 0$ el vector tx quedará representado por una flecha que tiene la misma dirección de la flecha que representa a x pero su longitud será t veces mayor. Si $t < 0$, el vector tx quedará representado por una flecha cuya dirección sea opuesta a la de x y con una longitud de $|t|$ veces la longitud de la flecha que representa a x. Dos vectores no nulos x y y se denominan *paralelos* si $y = tx$ para cualquier número real t no nulo. (Así, los vectores no nulos con direcciones iguales u opuestas, son paralelos.)

Para describir algebraicamente la multiplicación por escalares, introdúzcase de nuevo un sistema de coordenadas en un plano que contenga al vector x tal que x parta del origen. Si el extremo de x tiene por coordenadas a (a_1, a_2), entonces puede mostrarse fácilmente que las coordenadas del extremo de tx son (ta_1, ta_2). (Véase el ejercicio 5.)

Las descripciones algebraicas de la suma de vectores y de la multiplicación de vectores por escalares en un plano, implican las siguientes propiedades para vectores arbitrarios x, y, y z y números reales arbitrarios a y b:

1. $x + y = y + x$.
2. $(x + y) + z = x + (y + z)$.
3. Existe un vector llamado 0 tal que $x + 0 = x$ para todo vector x.
4. Para cada vector x existe un vector y tal que $x + y = 0$.
5. $1x = x$.
6. $(ab)x = a(bx)$.
7. $a(x + y) = ax + ay$.
8. $(a + b)x = ax + bx$.

Argumentos semejantes a los antes mencionados muestran que estas 8 propiedades, así como las interpretaciones geométricas de suma de vectores y multiplicación por escalares, son válidas para vectores que actúan en el espacio y no sólo en un plano. Utilizaremos estos resultados para escribir las ecuaciones de rectas y planos en el espacio.

Considérese primero la ecuación de una recta en el espacio que pasa por dos puntos distintos P y Q. Sea O el origen de un sistema de coordenadas en el espacio y sean u y v los vectores que parten de O y terminan respectivamente en P y Q. Si w es el vector que principia en P y termina en Q, la suma "cabeza con cola" muestra que $u + w = v$ y por tanto

$w = v - u$ donde $-u$ representa al vector $(-1)u$. (Véase la fig. 1.3 en donde el cuadrilátero $OPQR$ es un paralelogramo.) Como un múltiplo de escalar w es paralelo a w, pero posiblemente de una longitud diferente a w, cualquier punto de la recta que une a P y Q se puede obtener como el extremo del vector que principia en P y que tiene la forma tw para

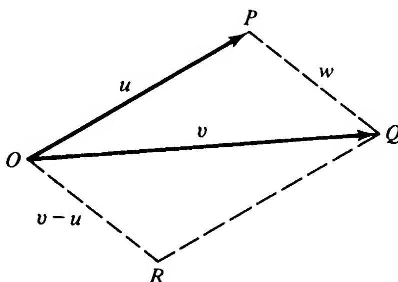


figura 1.3

algún número real t. Recíprocamente, el extremo de cada vector de la forma tw que principia en P yace en la línea que une a P y Q. Luego, una ecuación de la recta que pasa por P y Q es $x = u + tw = u + t(v-u)$, donde t es un número real y x es un punto arbitrario de la recta. Véase también que el extremo R del vector $v - u$ de la fig. 1.3 tiene coordenadas iguales a la diferencia de las coordenadas de Q y P.

Ejemplo. Encontremos la ecuación de la recta que pasa por los puntos P y Q de coordenadas $(-2, 0, 1)$ y $(4, 5, 3)$, respectivamente. El extremo R del vector que parte del origen y que tiene la misma dirección que el vector que principia en P y termina en Q, tiene como coordenadas $(4, 5, 3) - (-2, 0, 1) = (6, 5, 2)$. Luego, la ecuación buscada será:

$$x = (-2, 0, 1) + t(6, 5, 2).$$

Ahora, sean P, Q y R tres puntos no colineales en el espacio. Estos puntos determinan un plano único cuya ecuación puede ser encontrada mediante el uso de nuestras anteriores observaciones sobre vectores. Sean u y v los vectores que parten de P y terminan, respectivamente, en Q y R. Obsérvese que cualquier punto del plano que contenga a P, Q y R es el extremo S de un vector x que principia en P y tiene la forma $t_1u + t_2v$ para cualquier par de números reales t_1 y t_2. El extremo de t_1u será el punto de intersección de la recta que pasa por P y Q con la recta que pasa por S y es paralela a la recta que pasa por P y R. (Véase fig. 1.4.) Un procedimiento análogo permitirá localizar t_2v.

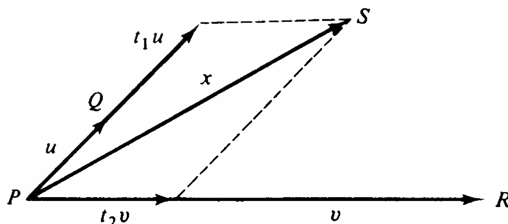


figura 1.4

Mas aún, para cualquier par de números reales t_1 y t_2, $t_1u + t_2v$ es un vector ubicado en el plano que contiene a P, Q y R. Por lo tanto, la ecuación del plano que contiene a P, Q y R es

$$x = P + t_1u + t_2v,$$

donde t_1 y t_2 son números reales arbitrarios y x es un punto cualquiera del plano.

Ejemplo. Sean P, Q y R puntos de coordenadas $(1, 0, 2)$, $(-3, -2, 4)$ y $(1, 8, -5)$, respectivamente. El extremo del vector que parte del origen y tiene la misma longitud y dirección que el vector que va de P a Q es $(-3, -2, 4) - (1, 0, 2) = (-4, -2, 2)$; de la misma forma, el extremo del vector que parte del origen y tiene la misma longitud y dirección que el vector que va de P a R es $(1, 8, -5) - (1, 0, 2) =$ $= (0, 8, -7)$. Luego, la ecuación del plano que contiene a los tres puntos dados es

$$x = (1, 0, 2) + t_1(-4, -2, 2) + t_2(0, 8, -7).$$

Cualquier estructura matemática que posea las ocho propiedades de la página 3 se llama "espacio vectorial". En la sección siguiente definiremos formalmente un espacio vectorial y consideraremos muchos ejemplos de espacios vectoriales distintos a los antes mencionados.

EJERCICIOS

1. Determinar si los vectores que parten del origen y terminan en los siguientes pares de puntos son paralelos.

   (a) $(3, 1, 2)$ y $(6, 4, 2)$
   (b) $(-3, 1, 7)$ y $(9, -3, -21)$
   (c) $(5, -6, 7)$ y $(-5, 6, -7)$
   (d) $(2, 0, -5)$ y $(5, 0, -2)$

2. Encontrar las ecuaciones de las rectas que pasan por los siguientes pares de puntos en el espacio.

(a) (3, −2, 4) y (−5, 7, 1)
(b) (2, 4, 0) y (−3, −6, 0)
(c) (3, 7, 2) y (3, 7, −8)
(d) (−2, −1, 5) y (3, 9, 7)

3. Encontrar las ecuaciones de los planos que contienen los siguientes puntos en el espacio.

(a) (2, −5, −1), (0, 4, 6) y (−3, 7, 1)
(b) (3, −6, 7), (−2, 0, −4) y (5, −9, −2)
(c) (−8, 2, 0), (1, 3, 0) y (6, −5, 0)
(d) (1, 1, 1), (5, 5, 5) y (−6, 4, 2)

4. ¿Cuáles son las coordenadas del vector *0* en el plano Euclidiano que satisfacen la condición 3 de la página 3. Demostrar que esta selección de coordenadas satisface la condición 3.

5. Demostrar que si el vector x parte del origen del plano Euclidiano y termina en el punto de coordenadas (a_1, a_2), entonces el vector tx que parte del origen termina en el punto de coordenadas (ta_1, ta_2).

6. Demostrar que las diagonales de un paralelogramo se bisectan.

1.2 ESPACIOS VECTORIALES

Debido a que entidades tan diversas como las fuerzas que operan en un plano y los polinomios con coeficientes reales permiten definiciones naturales de suma y multiplicación por escalares que poseen las propiedades 1 a 8 de la página 3, es evidente que se deban abstraer dichas propiedades en la siguiente definición.

DEFINICIÓN. Un espacio vectorial *(o* espacio lineal) V sobre un campo * F *consiste de un conjunto en el que están definidas dos operaciones (llamadas* adición *y* multiplicación por escalares, *respectivamente), tal que para cualquier par de elementos* x *y* y *en* V *existe un elemento único* x + y *en* V, *y para cada elemento* a *en* F *y cada elemento* x *en* V *exista un elemento único* ax *en* V, *de manera que se cumplan las siguientes condiciones:*

* Ver apéndice C. Sin embargo, con muy pocas excepciones, el lector puede interpretar la palabra "campo" como "campo de los números reales" (denotado por R) o "campo de los números complejos" (denotado por C).

(VS 1) *Para toda* x, y *en* V, x + y = y + x *(conmutatividad de la adición).*

(VS 2) *Para toda* x, y, z *en* V, (x + y) + z = x + (y + z) *(asociatividad de la adición).*

(VS 3) *Existe un elemento en* V *llamado* 0 *tal que* x + 0 = x *para toda* x *en* V.

(VS 4) *Para cada elemento* x *en* V, *existe un elemento* y *en* V *tal que* x + y = 0.

(VS 5) *Para cada elemento* x *en* V, 1x = x.

(VS 6) *Para cada par* a, b *de elementos en* F *y cada elemento* x *en* V, (ab)x = a(bx).

(VS 7) *Para cada elemento* a *en* F *y cada par de elementos* x, y *en* V, a(x + y) = ax + ay.

(VS 8) *Para cada par de elementos* a, b *en* F *y cada elemento* x *en* V, (a + b)x = ax + bx.

Los elementos x + y *y* ax *se denominan, respectivamente,* suma *de* x *y* y *y el* producto *de* a *y* x.

Los elementos del campo *F* se llaman *escalares* y los elementos del espacio vectorial V se llaman *vectores.* El lector no debe confundir este uso de la palabra "vector" con la entidad física tratada en la sección 1.1; ahora, la palabra "vector" se utilizará para describir cualquier elemento de un espacio vectorial.

Frecuentemente, un espacio vectorial será tratado en el texto sin mencionar explícitamente su campo de escalares. El lector cuidará de recordar, sin embargo, que todo espacio vectorial debe considerarse como un espacio vectorial sobre un campo, el que se denotará por *F*.

En el resto de la sección introduciremos diversos ejemplos importantes de espacios vectoriales que serán estudiados a través del texto. Obsérvese que al describir un espacio vectorial no sólo es necesario especificar los vectores, también las operaciones de suma y multiplicación por escalares.

Un objeto de la forma $(a_1, \ldots, a_n)$, donde los valores o entradas a_i son elementos de un campo *F*, se denomina *n-dimensional* * con valores de *F*. Dos *n*-dimensionales $(a_1, \ldots, a_n)$ y $(b_1, \ldots, b_n)$ se definen como iguales si y sólo si $a_i = b_i$ para $i = 1, 2, \ldots, n$.

Ejemplo 1. El espacio vectorial F^n de *n*-dimensionales con valores de un campo *F*.

El conjunto de todas las *n*-dimensionales con valores de un campo *F* forma un espacio vectorial, que denotaremos por F^n, bajo las operaciones de suma y multiplicación coordinada (elemento a elemento); esto es, si $x = (a_1, \ldots, a_n) \in \mathsf{F}^n$, $y = (b_1, \ldots, b_n) \in \mathsf{F}^n$, y $c \in F$, entonces

$$x + y = (a_1 + b_1, \ldots, a_n + b_n) \quad \text{y} \quad cx = (ca_1, \ldots, ca_n).$$

* N. del T. En algunos libros de álgebra lineal a los *n*-dimensionales se les da el nombre de *n*-adas, *n*-uplas, *n*-tuplas y otros más, pero aquí preferiremos la citada denominación.

Por ejemplo, en R^4

$$(3, -2, 0, 5) + (-1, 1, 4, 2) = (2, -1, 4, 7)$$

y

$$-5(1, -2, 0, 3) = (-5, 10, 0, -15).$$

Los elementos de F^n a menudo se escribirán como *vectores columna:*

$$\begin{pmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{pmatrix}$$

en vez de como *vectores renglón* $(a_1, \ldots, a_n)$. Puesto que un 1-dimensional con valor de F puede ser visto como elemento de F, escribiremos F en vez de F^1 para el espacio vectorial de los 1-dimensionales de F.

Una *matriz* de $m \times n$ con valores de un campo F es un arreglo rectangular de la forma

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

donde cada elemento $a_{ij} (1 \leq i \leq m,\ 1 \leq j \leq n)$ pertenece a F. Los elementos $a_{i1}, a_{i2}, \ldots, a_{in}$ de la matriz anterior forman el *i-ésimo renglón* de la matriz y se considerarán a menudo como un vector renglón en F^n, mientras que los elementos $a_{1j}, a_{2j}, \ldots, a_{mj}$ forman la *columna j-ésima* de la matriz y serán a menudo considerados como un vector columna en F^m. La matriz de $m \times n$ en la que cada elemento es igual a 0 se denomina *matriz cero.*

En este libro escribiremos las matrices con letras mayúsculas cursivas (p. ej., A, B y C) y denotaremos al elemento de la matriz A ubicado en el renglón i y la columna j por A_{ij}. Además, si el número de renglones es igual al número de columnas de una matriz, ésta se denominará *cuadrada.*

Dos matrices de $m \times n$, A y B se definen como iguales si y sólo si sus elementos correspondientes son iguales; esto es, si y sólo si $A_{ij} = B_{ij}$ para $1 \leq i \leq m$ y $1 \leq j \leq n$.

Ejemplo 2. El espacio vectorial $\mathsf{M}_{m\times n}(F)$ de matrices $m \times n$ con valores de un campo F.

El conjunto de todas las matrices de $m \times n$ con elementos de un campo F es un espacio vectorial, que denotaremos por $\mathsf{M}_{m\times n}(F)$, bajo las siguientes operaciones de suma y multiplicación por escalares: para A, $B \in \mathsf{M}_{m\times n}(F)$ y $c \in F$,

$$(A + B)_{ij} = A_{ij} + B_{ij} \quad \text{y} \quad (cA)_{ij} = cA_{ij}.$$

Por ejemplo

$$\begin{pmatrix} 2 & 0 & -1 \\ 1 & -3 & 4 \end{pmatrix} + \begin{pmatrix} -5 & -2 & 6 \\ 3 & 4 & -1 \end{pmatrix} = \begin{pmatrix} -3 & -2 & 5 \\ 4 & 1 & 3 \end{pmatrix}$$

y

$$-3\begin{pmatrix} 1 & 0 & -2 \\ -3 & 2 & 3 \end{pmatrix} = \begin{pmatrix} -3 & 0 & 6 \\ 9 & -6 & -9 \end{pmatrix}$$

en $\mathsf{M}_{2\times 3}(R)$.

Ejemplo 3. El espacio vectorial $\mathcal{F}(S, F)$ de todas las funciones de un conjunto S en un campo F.

Sea S un conjunto no vacío y F cualquier campo, y sea $\mathcal{F}(S, F)$ el conjunto de todas las funciones que van de S a F. Dos elementos f y g en $\mathcal{F}(S, F)$ se definen como iguales si $f(s) = g(s)$ para cada $s \in S$. El conjunto $\mathcal{F}(S, F)$ es un espacio vectorial bajo las operaciones de suma y multiplicación por escalares definidas para f, g, $\in \mathcal{F}(S, F)$ y $c \in F$ por

$$(f + g)(s) = f(s) + g(s) \quad \text{y} \quad (cf)(s) = c[f(s)]$$

para cada $s \in S$. Nótese que éstas son las operaciones normales de suma y producto por escalares utilizadas en cálculo.

Un *polinomio* con coeficientes de un campo F es una expresión de la forma

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0,$$

donde n es un entero no negativo y $a_n, \ldots, a_0$ son elementos de F. Si $f(x) = 0$, esto es, si $a_n = \ldots = a_0 = 0$, entonces $f(x)$ se llama el *polinomio cero* y se dice que el grado de $f(x)$ es -1; de otra forma, se define el *grado* de un polinomio como el mayor exponente de x que aparece en la representación

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0$$

correspondiente a un coeficiente no nulo. Nótese que los polinomios de grado cero son funciones de la forma $f(x) = c$ para algún escalar c no nulo.

Dos polinomios $f(x)$ y $g(x)$ son iguales si y sólo si tienen el mismo grado y los coeficientes de potencias iguales son iguales.

Cuando F es un campo que contiene un número infinito de elementos, normalmente consideraremos un polinomio con coeficientes de F como una función de F en F. En este caso, el valor de la función

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0$$

en $c \in F$ es el escalar

$$f(c) = a_n c^n + a_{n-1}c^{n-1} + \ldots + a_0.$$

Aquí, es posible utilizar cualquiera de las dos notaciones f o $f(x)$ para la función polinomial

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0.$$

Ejemplo 4. El espacio vectorial $\mathsf{P}(F)$ de todos los polinomios con coeficientes de un campo F.

El conjunto de todos los polinomios con coeficientes de un campo F es un espacio vectorial, que denotaremos por $\mathsf{P}(F)$, bajo las siguientes operaciones:

Para

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0$$

y

$$g(x) = b_n x^n + b_{n-1}x^{n-1} + \ldots + b_0$$

en $\mathsf{P}(F)$ y $c \in F$,

$$(f + g)(x) = (a_n + b_n)x^n + (a_{n-1} + b_{n-1})x^{n-1} + \ldots + (a_0 + b_0)$$

y

$$(cf)(x) = ca_n x^n + ca_{n-1}x^{n-1} + \ldots + ca_0.$$

Veremos en el ejercicio 21 de la sección 2.4 que el espacio vectorial que abajo se define es esencialmente el mismo que $\mathsf{P}(F)$.

Ejemplo 5. El espacio de todas las sucesiones finitas no nulas en un campo F.

Sea F cualquier campo. Una *sucesión* en F es una función σ de los enteros positivos en F. Como es usual, la sucesión σ tal que $\sigma(n) = a_n$, se escribirá como $\{a_n\}$. El espacio vectorial V de todas las sucesiones finitas no nulas en F está integrado por todas las sucesiones $\{a_n\}$ en F que solamente tienen un número finito de términos no nulos a_n. Si $\{a_n\}$ y $\{b_n\}$ son sucesiones en V y $t \in F$, entonces $\{a_n\} + \{b_n\}$ es aquella sucesión $\{c_n\}$ en V tal que $c_n = a_n + b_n (n = 1, 2, \ldots)$, y $t\{a_n\}$ es aquella sucesión $\{d_n\}$ en V tal que $d_n = ta_n (n = 1, 2, \ldots)$.

Nuestros dos ejemplos siguientes contienen conjuntos en los que están definidos una suma y un producto por escalares pero no se trata de espacios vectoriales.

Ejemplo 6. Sea $S = \{(a_1, a_2): a_1, a_2 \in R\}$. Para (a_1, a_2), $(b_1, b_2) \in S$ y $c \in R$, se definen

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 - b_2) \quad \text{y} \quad c(a_1, a_2) = (ca_1, ca_2).$$

Como (VS 1), (VS 2) y (VS 8) no se cumplen, S no es un espacio vectorial bajo estas operaciones.

Ejemplo 7. Sea S como en el ejemplo 6. Para (a_1, a_2), $(b_1, b_2) \in S$ y $c \in R$, definimos

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, 0) \quad \text{y} \quad c(a_1, a_2) = (ca_1, 0).$$

Luego, bajo estas operaciones, S no es un espacio vectorial pues (VS 3) (y por tanto (VS 4)) y (VS 5) fallan.

Esta sección concluirá con algunas de las consecuencias elementales de la definición de un espacio vectorial.

Teorema 1.1 *(Ley de cancelación para la suma vectorial). Si* x, y *y* z *son elementos de un espacio vectorial* V *tal que* x + z = y + z, *entonces* x = y.

DEMOSTRACIÓN. Existe un elemento v en V tal que $z + v = 0$ (VS 4). Luego, $x = x + 0 = x + (z + v) = (x + z) + v = (y + z) + v = y + (z + v) = y + 0 = y$ por (VS 2) y (VS 3). ■

Corolario 1. *El vector* 0 *descrito en (VS 3) es único.*

DEMOSTRACIÓN. Ejercicio.

Corolario 2. *El vector* y *descrito en (VS 4) es único.*

DEMOSTRACIÓN. Ejercicio.

El vector 0 descrito en (VS 3) se llama *vector cero* de V, y el vector y descrito en (VS 4) (esto es, el vector único tal que $x + y = 0$) se llama el *inverso aditivo* de x y se denota por $-x$.

El siguiente resultado contiene algunas de las propiedades elementales de la multiplicación por escalar.

Teorema 1.2 *En cualquier espacio vectorial* V, *son verdaderos los siguientes enunciados:*

(a) $0x = 0$ *para toda* $x \in V$.
(b) $(-a)x = -(ax)$ *para toda* $a \in F$ *y toda* $x \in V$.
(c) $a0 = 0$ *para toda* $a \in F$.

DEMOSTRACIÓN:

(a) Por (VS 8), (VS 1) y (VS 3) se tiene que

$$0x + 0x = (0 + 0)x = 0x = 0 + 0x.$$

Por tanto, $0x = 0$ por el Teorema 1.1.

(b) El elemento $-(ax)$ es el único elemento de V tal que $ax + [-(ax)] = 0$. Si $ax + (-a)x = 0$, el corolario 2 anterior implicaría que $(-a)x = -(ax)$. Pero por (VS 8), $ax + (-a)x = [a + (-a)]x = 0x$, y así $ax + (-a)x = 0x = 0$ por (a). Entonces, $(-a)x = -(ax)$.

La demostración de (c) es semejante a la demostración de (a). ■

EJERCICIOS

1. Determinar si las siguientes expresiones son falsas o verdaderas.

(a) Todo espacio vectorial contiene un vector cero.
(b) Un espacio vectorial puede tener más de un vector cero.
(c) En cualquier espacio vectorial $ax = bx$ implica que $a = b$.
(d) En cualquier espacio vectorial $ax = ay$ implica que $x = y$.
(e) Un elemento de F^n puede ser considerado como un elemento de $M_{n \times 1}(F)$.
(f) Una matriz de $m \times n$ tiene m columnas y n renglones.
(g) En $P(F)$ sólo se pueden sumar polinomios del mismo grado.
(h) Si f y g son polinomios de grado n, entonces $f + g$ es un polinomio de grado n.
(i) Si f es un polinomio de grado n y c es un escalar no nulo, entonces cf es un polinomio de grado n.
(j) Un elemento no nulo de F puede considerarse como un elemento de $P(F)$ de grado 0.

(k) Dos funciones en $\mathcal{F}(S, F)$ son iguales si y sólo si toman los mismos valores en cada punto de S.

2. Escribir el vector nulo de $\mathsf{M}_{3\times 4}(F)$.

3. Si

$$M = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix},$$

¿cuáles son M_{13}, M_{21} y M_{22}?

4. Realizar las operaciones indicadas.

(a) $\begin{pmatrix} 2 & 5 & -3 \\ 1 & 0 & 7 \end{pmatrix} + \begin{pmatrix} 4 & -2 & 5 \\ -5 & 3 & 2 \end{pmatrix}$

(b) $\begin{pmatrix} -6 & 4 \\ 3 & -2 \\ 1 & 8 \end{pmatrix} + \begin{pmatrix} 7 & -5 \\ 0 & -3 \\ 2 & 0 \end{pmatrix}$

(c) $4\begin{pmatrix} 2 & 5 & -3 \\ 1 & 0 & 7 \end{pmatrix}$

(d) $-5\begin{pmatrix} -6 & 4 \\ 3 & -2 \\ 1 & 8 \end{pmatrix}$

(e) $(2x^4 - 7x^3 + 4x + 3) + (8x^3 + 2x^2 - 6x + 7)$

(f) $(-3x^3 + 7x^2 + 8x - 6) + (2x^3 - 8x + 10)$

(g) $5(2x^7 - 6x^4 + 8x^2 - 3x)$

(h) $3(x^5 - 2x^3 + 4x + 2)$

Los Ejercicios 5 y 6 muestran por qué las definiciones de suma y multiplicación por escalares de matrices (como se definen en el ejemplo 2) son las adecuadas.

5. Richard Gard (Efectos de los castores en las truchas en Sagehen Creek, California. *J. Wildlife Management*, **25**, 221-242) reporta el siguiente número de truchas que atravesaron las represas de castores en Sagehen Creek:

Cruces a contracorriente

	Otoño	*Primavera*	*Verano*
Trucha arroyo	8	3	1
Trucha arcoiris	3	0	0
Trucha café	3	0	0

Cruces a favor de la corriente

	Otoño	*Primavera*	*Verano*
Trucha arroyo	9	1	4
Trucha arcoiris	3	0	0
Trucha café	1	1	0

Registrar los cruces a contracorriente y a favor de la corriente como datos en dos matrices de 3×3 y verificar que la suma de las dos matrices da el número total de cruces (a contracorriente y a favor) categorizada por especie de trucha y por estación.

6. Al final de mayo, un almacén de muebles tenía el siguiente inventario:

	Americano tradicional	*Español*	*Mediterráneo*	*Danés*
Conjuntos de sala	4	2	1	3
Conjuntos de alcoba	5	1	1	4
Conjuntos de comedor	3	1	2	6

Registrar estos datos como una matriz M de 3×4. Con el fin de prepararse para su venta de junio, el almacén decidió duplicar su inventario de cada uno de los rubros anteriores. Suponiendo que nada de la mercancía en inventario se vende hasta que los pedidos de muebles adicionales lleguen, se verifica que el inventario disponible después de recibir el pedido estará dado por la matriz $2M$. Si el inventario al final de junio queda dado por la matriz

$$A = \begin{pmatrix} 5 & 3 & 1 & 2 \\ 6 & 2 & 1 & 5 \\ 1 & 0 & 3 & 3 \end{pmatrix},$$

interpretar $2M - A$. ¿Cuántos conjuntos se vendieron durante la venta de junio?

7. Sea $S = \{0, 1\}$ y $F = R$, el campo de los números reales. En $\mathcal{F}(S, R)$, demostrar que $f = g$ y $f + g = h$ donde $f(x) = 2x + 1$, $g(x) = 1 + 4x - 2x^2$, y $h(x) = 5^x + 1$.

8. Demostrar que en cualquier espacio vectorial V, $(a + b)(x + y) = ax + ay + bx + by$ para toda $x, y \in$ V y cualquier $a, b \in F$.

9. Demostrar los Corolarios 1 y 2 del Teorema 1.1 y el Teorema 1.2(c).

10. Sea V el conjunto de todas las funciones diferenciales de valores reales definidas sobre la recta de los reales. Demostrar que V es un espacio vecto-

rial bajo las operaciones de suma y multiplicación por escalares definidas en el ejemplo 3.

11. Sea V = $\{0\}$ que conste de un único vector 0 y defínase $0 + 0 = 0$ y $c0 = 0$ para cada c de F. Demostrar que V es un espacio vectorial sobre F(V se llama el *espacio vectorial cero*).

12. Una función de valor real definida sobre la recta de los reales se llama *función par si* $f(-x) = f(x)$ para todo número real x. Demostrar que el conjunto de las funciones par definidas en la recta de los reales, con las operaciones de suma y multiplicación por escalares definidas en el ejemplo 3, es un espacio vectorial.

13. Sea V el conjunto de pares ordenados de números reales. Si (a_1, a_2) y (b_1, b_2) son elementos de V y c es un elemento de F, se definen

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2b_2) \quad \text{y} \quad c(a_1, a_2) = (ca_1, a_2).$$

¿Es V un espacio vectorial bajo esas operaciones? Verifique su respuesta.

14. Sea V = $\{(a_1, \ldots, a_n) \colon a_i \in C$ para $i = 1, 2, \ldots, n\}$. ¿Es V un espacio vectorial sobre el campo de los números reales con las operaciones de suma y multiplicación con correspondencia de elementos?

15. Sea V = $\{(a_1, \ldots, a_n) \colon a_i \in R$ para $i = 1, 2, \ldots, n\}$. ¿Es V un espacio vectorial sobre el campo de los números complejos bajo las operaciones de suma y multiplicación con correspondencia de elementos?

16. Sea V = $\{(a_1, a_2) \colon a_1, a_2 \in R\}$. Para (a_1, a_2), $(b_1, b_2) \in$ V y $c \in R$, defínase

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 + b_2)$$

y

$$c(a_1, a_2) = \begin{cases} (0, 0) & \text{si } c = 0 \\ \left(ca_1, \dfrac{a_2}{c}\right) & \text{si } c \neq 0. \end{cases}$$

¿Es V un espacio vectorial bajo estas operaciones? Justifique su respuesta.

17. Sea V = $\{(a_1, a_2) \colon \; a_1, a_2 \in C\}$. Para (a_1, a_2), $(b_1, b_2) \in$ V y $c \in C$, defínase

$$(a_1, a_2) + (b_1, b_2) = (a_1 + 2b_1, a_2 + 3b_2) \quad \text{y} \quad c(a_1, a_2) = (ca_1, ca_2).$$

¿Es V un espacio vectorial bajo estas operaciones? Justifique su respuesta.

18. Sea $V = \{(a_1, a_2): a_1, a_2 \in F\}$, donde F es un campo arbitrario. Defínase la suma de los elementos de V elemento a elemento, y para $c \in F$ y $(a_1, a_2) \in V$, defínase

$$c(a_1, a_2) = (a_1, 0).$$

¿Es V un espacio vectorial bajo estas operaciones? Justifique su respuesta.

1.3 SUBESPACIOS

Normalmente, en el estudio de cualquier estructura algebraica es interesante examinar subconjuntos que tengan la misma estructura que el conjunto que esté siendo considerado. Así, la noción apropiada de subestructura para espacios vectoriales se introduce en esta sección.

Definición. *Un subconjunto* W *de un espacio vectorial* V *sobre un campo* F *se llama un* subespacio *de* V *si* W *es un espacio vectorial sobre* F, *bajo las operaciones de suma y multiplicación por escalares definidas en* V.

En cualquier espacio vectorial V, es de hacer notar que V y $\{0\}$ son subespacios. Este último se denomina el *subespacio cero* de V.

Afortunadamente, no es necesario verificar todas las condiciones sobre espacios vectoriales con el objeto de demostrar que un subconjunto W de un espacio vectorial V es en realidad un subespacio. Como se sabe, las condiciones (VS 1), (VS 2), (VS 5), (VS 6), (VS 7) y (VS 8) se satisfacen para los elementos de V, las cuales, automáticamente se cumplen también para los elementos de un subconjunto V. Entonces, un subconjunto W de V es un subespacio de V si y sólo si las siguientes cuatro condiciones se satisfacen:

1. $x + y \in W$ siempre y cuando $x \in W$ y $y \in W$.
2. $ax \in W$ siempre que $a \in F$ y $x \in W$.
3. El vector cero de V pertenece a W.
4. El inverso aditivo de cada elemento de W pertenece a W.

En realidad, la condición 4 es redundante, como lo muestra el siguiente teorema.

Teorema 1.3 *Sea* V *un espacio vectorial y* W *un subconjunto de* V. *Entonces,* W *es un subespacio de* V *si y sólo si se satisfacen las tres condiciones siguientes:*

(a) $0 \in W$.
(b) $x + y \in W$ *siempre que* $x \in W$ *y* $y \in W$.
(c) $ax \in W$ *siempre que* $a \in F$ *y* $x \in W$.

DEMOSTRACIÓN. Si W es un subespacio de V, entonces W es un espacio vectorial bajo las operaciones de suma y multiplicación por escalares definidas en V. Tenemos entonces que se cumplen las condiciones (b) y (c), y existe un elemento $0' \in W$ tal que $x + 0' = x$ para toda $x \in W$. Pero también $x + 0 = x$, y por tanto $0' = 0$ por el Teorema 1.1. Luego entonces, también se satisface la condición (a).

Recíprocamente, si se satisfacen las condiciones (a), (b) y (c), la exposición que precede a este teorema muestra que W puede ser un subespacio de V si el inverso aditivo de cada elemento de W pertenece a W. Pero si $x \in W$, entonces $(-1)x$ pertenece a W por la condición (c), y $-x = (-1)x$ por el Teorema 1.2. De aquí que W sea un subespacio de V. ■

El teorema anterior proporciona un método sencillo para determinar si un subconjunto dado de un espacio vectorial es o no realmente un subespacio. En general, este resultado es el que se emplea para demostrar que un cierto subconjunto es un subespacio.

La *transpuesta* M^t de una matriz M de $m \times n$ es la matriz de $n \times m$ obtenida a partir de M mediante el intercambio de renglones con columnas; esto es $(M^t)_{ij} = M_{ji}$. Por ejemplo,

$$\begin{pmatrix} 1 & -2 & 3 \\ 0 & 5 & -1 \end{pmatrix}^t = \begin{pmatrix} 1 & 0 \\ -2 & 5 \\ 3 & -1 \end{pmatrix}.$$

Una *matriz simétrica* es una matriz M tal que $M^t = M$. Evidentemente, una matriz simétrica debe ser cuadrada. El conjunto W de todas las matrices simétricas en $M_{n \times n}(F)$ es un subespacio de $M_{n \times n}(F)$ ya que se satisfacen las condiciones del Teorema 1.3:

(a) La matriz cero es igual a su transpuesta y, por tanto, pertenece a W.

Puede probarse fácilmente que para matrices A y B y para escalares a y b cualesquiera, $(aA + bB)^t = aA^t + bB^t$. (Ver el ejercicio 3.) Usando este hecho, se pueden establecer fácilmente las condiciones (b) y (c) del Teorema 1.3 de la manera siguiente:

(b) Si $A \in W$ y $B \in W$, entonces $A = A^t$ y $B = B^t$. Ahora bien, $(A + B)^t = A^t + B^t = A + B$, de manera que $A + B \in W$.

(c) Si $A \in W$, entonces $A^t = A$. Luego, para toda $a \in F$, $(aA)^t = aA^t = aA$. Y así $aA \in W$.

Los siguientes ejemplos proporcionan más ilustraciones del concepto de subespacio. Los primeros tres son particularmente importantes.

Ejemplo 8. Las matrices diagonales en $\mathsf{M}_{n\times n}(F)$.

Sea M una matriz de $n \times n$. La *diagonal (principal)* de M consta de los términos M_{11}, M_{22}, . . . , M_{nn}. Una matriz D de $n \times n$ se llama *matriz diagonal* si todos los valores que no se encuentren sobre la diagonal de D son nulos, esto es, si $D_{ij} = 0$ para toda $i \neq j$. El conjunto de todas las matrices diagonales en $\mathsf{M}_{n\times n}(F)$ es un subespacio de $\mathsf{M}_{n\times n}(F)$.

Ejemplo 9. Los polinomios de grado menor o igual a n.

Sea n un entero no negativo y sea $\mathsf{P}_n(F)$ un conjunto que consista de todos los polinomios en $\mathsf{P}(F)$ que tengan grado menor o igual a n. (Nótese que el polinomio nulo es un elemento de $\mathsf{P}_n(F)$ pues su grado es -1.) Luego entonces, $\mathsf{P}_n(F)$ es un subespacio de $\mathsf{P}(F)$.

Ejemplo 10. Las funciones continuas de valores reales definidas en el eje de los reales R.

El conjunto $\mathsf{C}(R)$ formado por todas las funciones continuas de valor real definidas en R es un subespacio de $\mathcal{F}(R, R)$, donde $\mathcal{F}(R, R)$ es tal como se definió en el ejemplo 3.

Ejemplo 11. La *traza* de una matriz M de $n \times n$, denotada por $\text{tr}(M)$, es la suma de los valores de M ubicados en la diagonal; esto es, $\text{tr}(M) = M_{11} + M_{22} + \ldots + M_{nn}$. El conjunto de todas las matrices de $n \times n$ que tienen una traza igual a cero es un subespacio de $\mathsf{M}_{n\times n}(F)$. (Ver el ejercicio 6.)

Ejemplo 12. El conjunto de matrices en $\mathsf{M}_{m\times n}(F)$ que únicamente tengan elementos no negativos no es un subespacio de $\mathsf{M}_{m\times n}(F)$ ya que no se cumple la condición (c) del Teorema 1.3.

Los dos teoremas siguientes proporcionan métodos para formar subespacios a partir de otros subespacios.

Teorema 1.4 *Cualquier intersección de subespacios de un espacio vectorial* V *es un subespacio de* V.

DEMOSTRACIÓN. Sea $\mathcal{C}$ un conjunto de subespacios de V y sea W la intersección de todos los subespacios en $\mathcal{C}$. Como cada uno de los subespacios contiene al vector cero, $0 \in \mathsf{W}$. Sean $a \in F$ y x, y elementos de W; entonces x y y son elementos de cada subespacio en $\mathcal{C}$. De aquí concluimos que $x + y$ y ax son elementos de cada subespacio en $\mathcal{C}$ (porque la suma de vectores en un subespacio y el producto de un escalar y un vector del subespacio, ambos pertenecen a ese subespacio). Entonces $x + y \in \mathsf{W}$ y $ax \in \mathsf{W}$; luego entonces W es un subespacio de acuerdo con el Teorema 1.3. ■

Habiendo demostrado que la intersección de subespacios es un subespacio es lógico considerar la cuestión de si la unión de subespacios es o no un subespacio. Se puede ver fácilmente que la unión de subespacios debe satisfacer las condiciones (a) y (c) del Teorema 1.3 pero no necesariamente satisface la condición (b). De hecho, se puede demostrar de inmediato (ver ejercicio 18) que la unión de dos subespacios es un subespacio si y sólo si uno de los subespacios es un subconjunto de otro. Es normal, sin embargo, pensar que debería de existir un método para combinar ambos subespacios W_1 y W_2 para obtener un subespacio mayor (o sea, uno que contenga a W_1 y a W_2). Como sugerimos anteriormente, la clave para encontrar tal subespacio es la condición (b) del Teorema 1.3. Esta observación sugiere que debiéramos considerar la "suma" de dos subespacios (como se define a continuación).

Definición. *Si S_1 y S_2 son dos subconjuntos no vacíos de un espacio vectorial V, entonces la suma de S_1 y S_2, que se expresa como $S_1 + S_2$, es el conjunto $\{x + y: x \in S_1 \text{ y } y \in S_2\}$. La* suma *de cualquier número finito de subconjuntos no vacíos de V, $S_1, \ldots, S_n$, se define análogamente como el conjunto*

$$S_1 + \ldots + S_n = \{x_1 + \ldots + x_n: \quad x_i \in S_i \text{ para } i = 1, 2, \ldots, n\}.$$

Teorema 1.5 *Si W_1 y W_2 son subespacios de un espacio vectorial V, entonces $W_1 + W_2$ es un subespacio de V.*

DEMOSTRACIÓN. Sean W_1 y W_2 subespacios de V. Como $0 \in W_1$ y $0 \in W_2$, $0 = 0 + 0 \in W_1 + W_2$. Sea $a \in F$ y $x, y \in W_1 + W_2$; entonces existirá $x_1, y_1 \in W_1$ y $x_2, y_2 \in W_2$ tales que $x = x_1 + x_2$ y $y = y_1 + y_2$. Ahora bien,

$$x + y = (x_1 + x_2) + (y_1 + y_2) = (x_1 + y_1) + (x_2 + y_2)$$

es un elemento de $W_1 + W_2$ ya que $x_1 + y_1 \in W_1$ y $x_2 + y_2 \in W_2$, y

$$ax = a(x_1 + x_2) = ax_1 + ax_2$$

es un elemento de $W_1 + W_2$ ya que $ax_1 \in W_1$ y $ax_2 \in W_2$. Luego entonces $W_1 + W_2$ es, por el Teorema 1.3, un subespacio de V. ■

Corolario. *La suma de cualquier número finito de subespacios de V es un subespacio de V.*

Una clase especial de suma jugará un papel importante en los capítulos siguientes. Introduciremos un caso especial de este concepto en la siguiente definición.

Definición. *Se dice que un espacio vectorial* V *es la suma directa de* W_1 *y* W_2, *expresada como* $V = W_1 \oplus W_2$, *si* W_1 *y* W_2 *son subespacios de* V *tales que* $W_1 \cap W_2 = \{0\}$ *y* $W_1 + W_2 = V$.

Ejemplo 13. Sea $W_1 = \{(a, 0): a \in F\}$ y $W_2 = \{(0, b): b \in F\}$. Luego, $F^2 = W_1 \oplus W_2$.

Ejemplo 14. Una función g de valor real definida en R se llama *función par* si $g(-x) = g(x)$ para toda $x \in R$ y se llama *función impar* si $g(-x) = -g(x)$ para toda $x \in R$. Sean W_1 y W_2, respectivamente, los conjuntos de todas las funciones pares e impares en $\mathcal{F}(R, R)$.

Demostraremos que $\mathcal{F}(R, R) = W_1 \oplus W_2$. Puede verse fácilmente que W_1 y W_2 son subespacios de $\mathcal{F}(R, R)$. (Ver Ejercicio 19.) Supóngase que $g \in W_1 \cap W_2$; entonces g es al mismo tiempo una función par e impar. Así $g(-x) = g(x)$ y $g(-x) = -g(x)$ para cada $x \in R$ y, por lo tanto, g es la función cero. Por lo tanto, $W_1 \cap W_2 = \{0\}$. Sea $f \in \mathcal{F}(R, R)$, y defínase g, $h \in \mathcal{F}(R, R)$ como $g(x) = \frac{1}{2}[f(x) + f(-x)]$ y $h(x) = \frac{1}{2}[f(x) - f(-x)]$. Entonces g es una función par y h es una función impar tales que $f = g + h$. De aquí que $f \in W_1 + W_2$. Como f es un elemento arbitrario de $\mathcal{F}(R, R)$, se tiene que $\mathcal{F}(R, R) = W_1 + W_2$. Esto es, $\mathcal{F}(R, R)$ es la suma directa de W_1 y W_2.

Si W_1 y W_2 son subespacios de un espacio vectorial V tales que $W_1 + W_2 = V$, entonces, cada elemento de V puede expresarse como la suma de un elemento x_1 en W_1 y un elemento x_2 en W_2. Es posible que puedan existir muchas representaciones semejantes, es decir, que x_1 y x_2 no sean únicas. Por ejemplo, si

$$W_1 = \{(a_1, a_2, a_3) \in F^3: \quad a_3 = 0\}$$

y

$$W_2 = \{(a_1, a_2, a_3) \in F^3: \quad a_1 = 0\},$$

claramente $W_1 + W_2 = F^3$. De hecho, para cada $c \in F$, $(b_1, b_2, b_3) = (b_1, b_2 + c, 0) + (0, -c, b_3)$ es una representación de (b_1, b_2, b_3) como la suma de un elemento $(b_1, b_2 + c, 0)$ en W_1 y un elemento $(0, -c, b_3)$ en W_2. Así, en este ejemplo la representación de los elementos de F^3 como las sumas de un elemento en W_1 y un elemento en W_2 no es única. Nuestro próximo resultado determina cuándo existe este tipo de unicidad.

Teorema 1.6 *Sean* W_1 *y* W_2 *subespacios de un espacio vectorial* V. *Entonces* V *es la suma directa de* W_1 *y* W_2 *si y sólo si cada elemento de* V *puede ser escrito de manera única como* $x_1 + x_2$, *donde* $x_1 \in W_1$ *y* $x_2 \in W_2$.

DEMOSTRACIÓN. Supóngase que $V = W_1 \oplus W_2$. Como $V = W_1 + W_2$, cada elemento de V puede ser expresado como la suma de vectores en W_1 y W_2. Supóngase que algún elemento z en V puede ser escrito como

$z = x_1 + x_2$ y también como $z = y_1 + y_2$, donde $x_1, y_1 \in W_1$ y $x_2, y_2 \in W_2$. Entonces, $x_1 + x_2 = y_1 + y_2$ y así $x_1 - y_1 = y_2 - x_2$. Ahora bien, $x_1 - y_1 \in W_1$ puesto que x_1 y y_1 son elementos de W_1, y análogamente $y_2 - x_2 \in W_2$. Pero como $x_1 - y_1 = y_2 - x_2$ se deduce que $x_1 - y_1 = y_2 - x_2 \in W_1 \cap W_2 = \{0\}$.

Por lo tanto, $x_1 - y_1 = y_2 - x_2 = 0$, y así $x_1 = y_1$ y $x_2 = y_2$, lo que demuestra la unicidad de la representación de z como la suma de un elemento de W_1 y un elemento de W_2.

La demostración de la proposición recíproca se deja al lector como ejercicio. ■

EJERCICIOS

1. Decir si las siguientes proposiciones son verdaderas o falsas.

   (a) Si V es un espacio vectorial y W es subconjunto de V que es también un espacio vectorial, entonces W es un subespacio de V.
   (b) El conjunto vacío es un subespacio de todo espacio vectorial.
   (c) Si V es un espacio vectorial distinto del espacio vectorial cero $\{0\}$, entonces V contiene un subespacio W tal que $W \neq V$.
   (d) La suma de dos subconjuntos cualesquiera de V es un subespacio de V.
   (e) Una matriz diagonal $n \times n$ no puede tener más de n términos no nulos.
   (f) La traza de una matriz cuadrada es el producto de sus términos que se encuentran sobre la diagonal.

2. Determinar la transpuesta de cada una de las siguientes matrices. Además, si la matriz es cuadrada, calcular su traza.

   (a) $\begin{pmatrix} -4 & 2 \\ 5 & -1 \end{pmatrix}$ (b) $\begin{pmatrix} 0 & 8 & -6 \\ 3 & 4 & 7 \end{pmatrix}$

   (c) $\begin{pmatrix} -3 & 9 \\ 0 & -2 \\ 6 & 1 \end{pmatrix}$ (d) $\begin{pmatrix} 10 & 0 & -8 \\ 2 & -4 & 3 \\ -5 & 7 & 6 \end{pmatrix}$

   (e) $(1, \ -1, \ 3, \ 5)$ (f) $\begin{pmatrix} -2 & 5 & 1 & 4 \\ 7 & 0 & 1 & -6 \end{pmatrix}$

   (g) $\begin{pmatrix} 5 \\ 6 \\ 7 \end{pmatrix}$ (h) $\begin{pmatrix} -4 & 0 & 6 \\ 0 & 1 & -3 \\ 6 & -3 & 5 \end{pmatrix}$

3. Demostrar que $(aA + bB)^t = aA^t + bB^t$ para toda $A, B \in M_{m \times n}(F)$ y toda $a, b \in F$.

4. Demostrar que $(A^t)^t = A$ para toda $A \in \mathsf{M}_{m\times n}(F)$.

5. Demostrar que $A + A^t$ es simétrica para cualquier matriz cuadrada A.

6. Demostrar que $\mathrm{tr}(aA + bB) = a\,\mathrm{tr}(A) + b\,\mathrm{tr}(B)$ para toda A, $B \in \mathsf{M}_{n\times n}(F)$.

7. Demostrar que las matrices diagonales son matrices simétricas.

8. Verificar que los siguientes conjuntos son subespacios de R^3 bajo las operaciones de suma y multiplicación por escalares definidas en R^3.

(a) $\mathsf{W}_1 = \{(a_1, a_2, a_3) \in \mathsf{R}^3\colon\ a_1 = 3a_2 \text{ y } a_3 = -a_2\}$
(b) $\mathsf{W}_2 = \{(a_1, a_2, a_3) \in \mathsf{R}^3\colon\ 2a_1 + a_2 + 5a_3 = 0\}$
(c) $\mathsf{W}_3 = \{(a_1, a_2, a_3) \in \mathsf{R}^3\colon\ a_1 - 4a_2 - a_3 = 0\}$

9. Sean W_1, W_2, y W_3 como en el ejercicio 8. Describir $\mathsf{W}_1 \cap \mathsf{W}_2$, $\mathsf{W}_2 \cap \mathsf{W}_3$, y $\mathsf{W}_1 \cap \mathsf{W}_3$ y obsérvese que cada una es subespacio de R^3.

10. Verificar que $\mathsf{W}_1 = \{(a_1, \ldots, a_n) \in \mathsf{F}^n\colon\ a_1 + \ldots + a_n = 0\}$ es un subespacio de F^n pero que $\mathsf{W}_2 = \{(a_1, \ldots, a_n) \in \mathsf{F}^n\colon\ a_1 + \ldots + a_n = 1\}$ no lo es.

11. ¿Es el conjunto $\mathsf{W} = \{f \in \mathsf{P}(F)\colon\ f = 0 \text{ o } f \text{ tiene grado } n\}$ un subespacio de $\mathsf{P}(F)$ si $n \geq 1$? Justifique su respuesta.

12. Una matriz A de $m \times n$ se llama *triangular superior* si todos los términos ubicados por debajo de la diagonal valen cero, esto es, $A_{ij} = 0$ siempre que $i > j$. Verificar que las matrices triangulares superiores forman un subespacio de $\mathsf{M}_{m\times n}(F)$.

13. Verificar que para cualquier $s_0 \in S$, $\mathsf{W} = \{f \in \mathcal{F}(S, F)\colon\ f(s_0) = 0\}$ es un subespacio de $\mathcal{F}(S, F)$.

14. ¿Es el conjunto de todas las funciones diferenciales de valores reales definidas en R un subespacio de $\mathsf{C}(R)$? Justifique su respuesta.

15. Sea $\mathsf{C}^n(R)$ el conjunto de todas las funciones de valor real definidas en la recta de los reales que tiene una derivada n-ésima continua (y, por tanto, derivadas continuas de orden $1, 2, \ldots, n$). Verificar que $\mathsf{C}^n(R)$ es un subespacio de $\mathcal{F}(R, R)$.

16. Demostrar que un subconjunto W de un espacio vectorial V es un subespacio de V si y sólo si $\mathsf{W} \neq \varnothing$ y $ax \in \mathsf{W}$ y $x + y \in \mathsf{W}$ siempre que $a \in F$ y $x, y \in \mathsf{W}$.

17. Demostrar que un subconjunto W de un espacio vectorial V es un subespacio de V si y sólo si $\mathit{0} \in \mathsf{W}$ y $ax + y \in \mathsf{W}$ siempre que $a \in F$ y $x, y \in \mathsf{W}$.

18. Sean W_1 y W_2 subespacios de un espacio vectorial V. Demostrar que $W_1 \cup W_2$ es un subespacio de V si y sólo si $W_1 \subseteq W_2$ o $W_2 \subseteq W_1$.

19. Sean F_1 y F_2 campos. Una función $g \in \mathcal{F}(F_1, F_2)$ se llama *función par* si $g(-x) = g(x)$ para toda $x \in F_1$ y se llama *función impar* si $g(-x) = -g(x)$ para toda $x \in F_1$. Demostrar que el conjunto de todas las funciones pares en $\mathcal{F}(F_1, F_2)$ y el conjunto de todas las funciones impares en $\mathcal{F}(F_1, F_2)$ son subespacios de $\mathcal{F}(F_1, F_2)$.

20. Mostrar que F^n es la suma directa de los subespacios

$$W_1 = \{(a_1, \ldots, a_n) \in F^n: \quad a_n = 0\}$$

y

$$W_2 = \{(a_1, \ldots, a_n) \in F^n: \quad a_1 = \ldots = a_{n-1} = 0\}.$$

21. Sea W_1 el conjunto de polinomios f en $P(F)$ tales que $f(x) = 0$ o, en la representación

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0,$$

los coeficientes $a_0, a_2, a_4, \ldots$ de todas las potencias pares de x son iguales a cero. Análogamente, sea W_2 el conjunto de todos los polinomios g en $P(F)$ tales que $g(x) = 0$ o, en la representación

$$g(x) = b_m x^m + b_{m-1}x^{m-1} + \ldots + b_0,$$

los coeficientes $b_1, b_3, b_5, \ldots$ de todas las potencias impares de x son iguales a cero. Demostrar que $P(F) = W_1 \oplus W_2$.

22. Sea $W_1 = \{A \in M_{m\times n}(F): \quad A_{ij} = 0$ cuando $i > j\}$ y $W_2 = \{A \in M_{m\times n}(F): A_{ij} = 0$ cuando $i \leq j\}$. (W_1 es el conjunto de las matrices triangulares superiores definidas en el ejercicio 12.) Demostrar que $M_{m\times n}(F) = W_1 \oplus W_2$.

23. Sea V el espacio vectorial formado por todas las matrices triangulares superiores de $n \times n$ (como se definieron en el ejercicio 12), y sea W_1 el subespacio de V formado por todas las matrices diagonales. Demostrar que $V = W_1 \oplus W_2$, donde $W_2 = \{A \in V: \quad A_{ij} = 0$ cuando $i < j\}$.

24.* Demostrar que si W es un subespacio de V y $x_1, \ldots, x_n$ son elementos de W, entonces $a_1x_1 + \ldots + a_nx_n$ es un elemento de W para cualesquiera escalares $a_1, \ldots, a_n$ en F.

25. Una matriz M se llama *antisimétrica* si $M^t = -M$. Evidentemente una matriz antisimétrica es cuadrada. Demostrar que el conjunto de todas las matrices antisimétricas de $n \times n$ es un subespacio W_1 de $M_{n\times n}(R)$. Sea W_2 el subespacio de $M_{n\times n}(R)$ consistente de las matrices simétricas de $n \times n$. Demostrar que $M_{n\times n}(R) = W_1 \oplus W_2$.

---

* En otras secciones del libro haremos referencia a los problemas marcados con asterisco (*).

26. Sea $W_1 = \{A \in M_{n \times n}(F): A_{ij} = 0 \text{ cuando } i \leq j\}$ y sea W_2 el conjunto de matrices simétricas de $n \times n$. W_1 y W_2 son ambos subespacios de $M_{n \times n}(F)$. Demostrar que $M_{n \times n}(F) = W_1 \oplus W_2$. Compárense los ejercicios 25 y 26.

27. Demostrar el corolario del Teorema 1.5.

28. Completar la demostración del Teorema 1.6.

29. Sea W un subespacio de un espacio vectorial V sobre un campo F. Para toda $v \in V$ el conjunto $\{v\} + W = \{v + w: \; w \in W\}$ se llama *co-conjunto de* W *que contiene a* v. Es frecuente expresar este co-conjunto como $v + W$ en vez de $\{v\} + W$. Demostrar lo siguiente:

(a) $v + W$ es un subespacio de V si y sólo si $v \in W$.
(b) $v_1 + W = v_2 + W$ si y sólo si $v_1 - v_2 \in W$.

La suma y el producto por elementos de F puede definirse en el conjunto $S = \{v + W: \; v \in V\}$ de todos los co-conjuntos de W como sigue:

$$(v_1 + W) + (v_2 + W) = (v_1 + v_2) + W$$

para toda $v_1, v_2 \in V$ y

$$a(v + W) = av + W$$

para toda $v \in V$ y $a \in F$.

(c) Demostrar que las operaciones anteriores están bien definidas; es decir, mostrar que si $v_1 + W = v'_1 + W$ y $v_2 + W = v'_2 + W$, entonces

$$(v_1 + W) + (v_2 + W) = (v'_1 + W) + (v'_2 + W)$$

y

$$a(v_1 + W) = a(v'_1 + W)$$

para toda $a \in F$.

(d) Demostrar que el conjunto S es un espacio vectorial bajo las operaciones definidas anteriormente. Este espacio vectorial se llama *espacio cociente de* V *módulo* W y se expresa mediante V/W.

1.4 COMBINACIONES LINEALES Y SISTEMAS DE ECUACIONES LINEALES

En la sección 1.1 se mostró que la ecuación del plano que pasa por tres puntos no colineales P, Q y R en el espacio es $x = P + t_1u + t_2v$, donde u y v son los vectores que parten del origen y terminan, respectivamente,

en Q y R, y t_1 y t_2 son números reales cualesquiera. Un caso especial importante ocurre cuando P es el origen. En este caso la ecuación del plano se simplifica a $x = t_1u + t_2v$, y el conjunto de todos los puntos contenidos en este plano es un subespacio de R^3. (Esto se demostrará como el Teorema 1.7 de esta sección.) Expresiones de la forma $t_1u + t_2v$ donde t_1 y t_2 son escalares y u y v vectores, juegan un papel primordial en la teoría de los espacios vectoriales. La generalización apropiada de tales expresiones se expresa en la siguiente definición.

Definición. *Sea* V *un espacio vectorial y* S *un conjunto no vacío de* V. *Se dice que un vector* x *de* V *es una* combinación lineal de elementos de S, *si existe un número finito de elementos* $y_1, \ldots, y_n$ *en* S *y escalares* $a_1, \ldots, a_n$ *en* F *tales que* $x = a_1y_1 + \ldots + a_ny_n$. *En este caso, es común decir que* x *es una* combinación lineal *de* $y_1, \ldots, y_n$.

Obsérvese que en cualquier espacio vectorial V, $0x = 0$ para toda $x \in V$. Luego, el vector cero es una combinación lineal de cualquier subconjunto no vacío de V.

Ejemplo 15. La tabla 1.1 muestra el contenido vitamínico de 100 gramos de 12 alimentos con respecto a vitaminas A, B_1 (tiamina), B_2 (riboflavina), niacina y C (ácido ascórbico).

TABLA 1.1 Contenido de vitaminas de 100 gramos de algunos alimentos

	A (unidades)	B_1 *(mg)*	B_2 *(mg)*	*Niacina (mg)*	*C (mg)*
Compota de manzana	0	0.01	0.02	0.2	2
Manzanas frescas (recién cortadas)	90	0.03	0.02	0.1	4
Dulce relleno de coco y cubierto de chocolate	0	0.02	0.07	0.2	0
Almejas (únicamente la carne)	100	0.10	0.18	1.3	10
Pastel de molde de masa	0	0.05	0.06	0.3	0
Féculas cocidas (no enriquecidas)	(0)*	0.01	0.01	0.1	(0)
Jaleas y conservas	10	0.01	0.03	0.2	2
Tarta de natillas de coco (horneada con harina)	0	0.02	0.02	0.4	0
Arroz café crudo	(0)	0.34	0.05	4.7	(0)
Salsa de soya	0	0.02	0.25	0.4	0
Spaghetti horneado (no enriquecido)	0	0.01	0.01	0.3	0
Arroz silvestre crudo	(0)	0.45	0.63	6.2	(0)

* Los ceros entre paréntesis indican que la cantidad de vitamina presente es casi nula o demasiado pequeña para medirse.

FUENTE: Composición de alimentos (Manual de Agricultura Número 8) por Bernice K. Watt y Annabel L. Merrill. División de Investigación de la Economía y Alimentación del Consumidor, Departamento de Agricultura de los Estados Unidos, 1963.

Registraremos el contenido vitamínico de 100 gramos de cada alimento como un vector columna en R^5—por ejemplo, el vector vitamínico para la compota de manzana es

$$\begin{pmatrix} 0.00 \\ 0.01 \\ 0.02 \\ 0.20 \\ 2.00 \end{pmatrix}.$$

Considerando los vectores vitamínicos para el pastel de molde, la tarta de natillas de coco, el arroz café, la salsa de soya y el arroz silvestre, se ve que

$$\begin{pmatrix} 0.00 \\ 0.05 \\ 0.06 \\ 0.30 \\ 0.00 \end{pmatrix} + \begin{pmatrix} 0.00 \\ 0.02 \\ 0.02 \\ 0.40 \\ 0.00 \end{pmatrix} + \begin{pmatrix} 0.00 \\ 0.34 \\ 0.05 \\ 4.70 \\ 0.00 \end{pmatrix} + 2\begin{pmatrix} 0.00 \\ 0.02 \\ 0.25 \\ 0.40 \\ 0.00 \end{pmatrix} = \begin{pmatrix} 0.00 \\ 0.45 \\ 0.63 \\ 6.20 \\ 0.00 \end{pmatrix}.$$

Luego, el vector vitamínico para el arroz silvestre crudo es una combinación lineal de los vectores de vitaminas para el pastel de molde, tarta de natillas de coco, arroz café crudo y salsa de soya. Así, 100 gramos de pastel de molde, 100 gramos de tarta de natillas de coco, 100 gramos de arroz café crudo y 200 gramos de salsa de soya proporcionan exactamente las mismas cantidades de las 5 vitaminas que 100 gramos de arroz silvestre crudo. De una manera análoga, como

$$2\begin{pmatrix} 0.00 \\ 0.01 \\ 0.02 \\ 0.20 \\ 2.00 \end{pmatrix} + \begin{pmatrix} 90.00 \\ 0.03 \\ 0.02 \\ 0.10 \\ 4.00 \end{pmatrix} + \begin{pmatrix} 0.00 \\ 0.02 \\ 0.07 \\ 0.20 \\ 0.00 \end{pmatrix} + \begin{pmatrix} 0.00 \\ 0.01 \\ 0.01 \\ 0.10 \\ 0.00 \end{pmatrix} + \begin{pmatrix} 10.00 \\ 0.01 \\ 0.03 \\ 0.20 \\ 2.00 \end{pmatrix} + \begin{pmatrix} 0.00 \\ 0.01 \\ 0.01 \\ 0.30 \\ 0.00 \end{pmatrix} = \begin{pmatrix} 100.00 \\ 0.10 \\ 0.18 \\ 1.30 \\ 10.00 \end{pmatrix},$$

200 gramos de compota de manzana, 100 gramos de manzanas frescas, 100 gramos de dulce de chocolate, 100 gramos de féculas, 100 gramos de jalea y 100 gramos de spaghetti proporcionan exactamente las mismas cantidades de las 5 vitaminas que 100 gramos de almejas.

A través de los capítulos 1 y 2 se encontrarán muchas situaciones diferentes en las cuales será necesario determinar si un vector puede ser expresado como una combinación lineal de otros vectores. El cómo es posible hacerlo se reduce a un problema de solución de un sistema de

ecuaciones lineales. Ilustraremos esta importante técnica determinando si el vector (8, 15, 15, 12) en R^4 puede escribirse como una combinación lineal de $y_1 = (1, 2, 1, 2)$, $y_2 = (-2, -4, -2, -4)$, $y_3 = (1, 4, 2, 0)$, $y_4 = (2, 7, 5, 0)$, y $y_5 = (3, 7, 2, 6)$. Por tanto, deberemos determinar si existen escalares a_1, a_2, a_3, a_4 y a_5 tales que

$$\begin{aligned}(8,15, 15, 12) &= a_1y_1 + a_2y_2 + a_3y_3 + a_4y_4 + a_5y_5\\ &= a_1(1, 2, 1, 2) + a_2(-2, -4, -2, -4) + a_3(1, 4, 2, 0)\\ &\quad + a_4(2, 7, 5, 0) + a_5(3, 7, 2, 6)\\ &= (a_1 - 2a_2 + a_3 + 2a_4 + 3a_5, 2a_1 - 4a_2 + 4a_3\\ &\quad + 7a_4 + 7a_5, a_1 - 2a_2 + 2a_3 + 5a_4 + 2a_5,\\ &\quad 2a_1 - 4a_2 + 6a_5).\end{aligned}$$

Se puede ver ahora fácilmente que (8, 15, 15, 12) puede ser expresado como una combinación lineal de y_1, y_2, y_3, y_4 y y_5, si y sólo si existen escalares $(a_1, a_2, a_3, a_4, a_5)$ que satisfacen el sistema de ecuaciones lineales

$$\begin{cases} a_1 - 2a_2 + a_3 + 2a_4 + 3a_5 = 8\\ 2a_1 - 4a_2 + 4a_3 + 7a_4 + 7a_5 = 15\\ a_1 - 2a_2 + 2a_3 + 5a_4 + 2a_5 = 15\\ 2a_1 - 4a_2 \qquad\qquad\qquad + 6a_5 = 12 \end{cases} \qquad (1)$$

que se obtuvo igualando las coordenadas correspondientes de la ecuación anterior.

Para resolver el sistema de ecuaciones (1) se substituirá éste por otro que tenga las mismas soluciones pero que sea mucho más sencillo de resolver. El procedimiento que utilizaremos expresará algunas de las incógnitas en términos de otras eliminando algunas de ellas en todas las ecuaciones, menos en una. Para empezar, eliminemos a_1 de la segunda, tercera y cuarta ecuaciones del sistema (1). Esta eliminación puede realizarse sumando -2 veces la primera ecuación a la segunda, -1 vez la primera ecuación a la tercera y -2 veces la primera ecuación a la cuarta; el resultado será el nuevo sistema

$$\begin{cases} a_1 - 2a_2 + a_3 + 2a_4 + 3a_5 = 8\\ \qquad\qquad 2a_3 + 3a_4 + a_5 = -1\\ \qquad\qquad a_3 + 3a_4 - a_5 = 7\\ \qquad\quad -2a_3 - 4a_4 \qquad = -4 \end{cases} \qquad (2)$$

en el cual se han eliminado a_1 y a_2 en todas las ecuaciones, excepto en la primera. Continuando con el sistema de ecuaciones (2), agregaremos múltiplos de la segunda ecuación a las otras con objeto de eliminar a_3 de las ecuaciones menos en la segunda. En este caso debemos sumar $-\frac{1}{2}$

veces la segunda ecuación a la primera para eliminar a_3 de ésta. Nótese, sin embargo, que si se intercambian la segunda y la tercera ecuaciones, los cálculos necesarios se simplifican. Entonces, intercambiaremos la segunda y tercera ecuaciones del sistema (2) para obtener

$$\begin{cases} a_1 - 2a_2 + a_3 + 2a_4 + 3a_5 = 8 \\ a_3 + 3a_4 - a_5 = 7 \\ 2a_3 + 3a_4 + a_5 = -1 \\ -2a_3 - 4a_4 = -4. \end{cases} \tag{3}$$

Ahora, sumando -1 veces la segunda ecuación a la primera, -2 veces la segunda ecuación a la tercera y 2 veces la segunda ecuación a la cuarta, el sistema de ecuaciones (3) se transforma en

$$\begin{cases} a_1 - 2a_2 - a_4 + 4a_5 = 1 \\ a_3 + 3a_4 - a_5 = 7 \\ -3a_4 + 3a_5 = -15 \\ 2a_4 - 2a_5 = 10. \end{cases} \tag{4}$$

A continuación debemos sumar múltiplos de la tercera ecuación a las otras con objeto de eliminar a_4 en cada una de las ecuaciones del sistema (4), excepto en la tercera. De nuevo, los cálculos se simplifican si se realiza una operación preliminar—multiplicar la tercera ecuación por $-\frac{1}{3}$. Esto da

$$\begin{cases} a_1 - 2a_2 - a_4 + 4a_5 = 1 \\ a_3 + 3a_4 - a_5 = 7 \\ a_4 - a_5 = 5 \\ 2a_4 - 2a_5 = 10. \end{cases} \tag{5}$$

Por último, en el sistema (5) añadamos 1 vez la tercera ecuación a la primera, -3 veces la tercera ecuación a la segunda y -2 veces la tercera ecuación a la cuarta para obtener

$$\begin{cases} a_1 - 2a_2 + 3a_5 = 6 \\ a_3 + 2a_5 = -8 \\ a_4 - a_5 = 5 \\ 0 = 0. \end{cases} \tag{6}$$

El sistema de ecuaciones (6) es un sistema de la forma deseada: es fácil de resolver para a_1, a_3 y a_4 (las incógnitas que aparecen como primera incógnita presente en alguna de las ecuaciones) en términos de otras incógnitas (a_2 y a_5). Escribiendo de nuevo el sistema (6), encontramos que

$$\begin{aligned} a_1 &= 2a_2 - 3a_5 + 6 \\ a_3 &= \qquad\; -2a_5 - 8 \\ a_4 &= \qquad\qquad a_5 + 5 \end{aligned}$$

Entonces, para cualquier selección de los escalares a_2 y a_5, un vector de la forma

$$\begin{aligned} &(a_1, a_2, a_3, a_4, a_5) \\ &\quad = (2a_2 - 3a_5 + 6,\ a_2,\ -2a_5 - 8,\ a_5 + 5,\ a_5) \\ &\quad = a_2(2, 1, 0, 0, 0) + a_5(-3, 0, -2, 1, 1) + (6, 0, -8, 5, 0) \end{aligned}$$

será solución del sistema original de ecuaciones (1). En particular, el vector (6, 0. −8, 5, 0) obtenido al hacer $a_2 = 0$ y $a_5 = 0$ es una solución del sistema (1). Entonces,

$$(8, 15, 15, 12) = 6y_1 + 0y_2 - 8y_3 + 5y_4 + 0y_5,$$

de manera que (8, 15, 15, 12) es una combinación lineal de y_1, y_2, y_3, y_4 y y_5.

El procedimiento que acabamos de ilustrar puede utilizarse para resolver cualquier sistema de ecuaciones lineales. Obsérvese que se utilizaron tres tipos de operaciones para resolver el sistema original.

1. Intercambio del orden de cualquier par de ecuaciones en el sistema.
2. Multiplicación de cualquier ecuación por una constante *no nula.*
3. Suma de cualquier múltiplo constante de una ecuación a otra.

Estas operaciones se utilizaron hasta obtener un sistema de ecuaciones con las siguientes propiedades:

1. El primer coeficiente no nulo de cualquier ecuación es uno.
2. Si una incógnita es la primera con coeficiente no nulo en alguna ecuación, entonces dicha incógnita aparece con un coeficiente nulo en cada una de las otras ecuaciones.
3. La primera incógnita con coeficiente no nulo en cualquier ecuación tiene subíndice mayor que el de la primera incógnita con coeficiente no nulo en cualquier ecuación precedente.

Para ayudar a aclarar el significado de estas propiedades, nótese que ninguno de los siguientes sistemas satisface estas condiciones.

$$\begin{cases} x_1 + 3x_2 \quad\quad + \; x_4 = \;\; 7 \\ \quad\quad\quad\quad 2x_3 - 5x_4 = -1 \end{cases} \tag{7}$$

$$\begin{cases} x_1 - 2x_2 + 3x_3 \quad + \; x_5 = -5 \\ \quad\quad\quad\quad x_3 \quad - 2x_5 = \;\; 9 \\ \quad\quad\quad\quad\quad x_4 + 3x_5 = \;\; 6 \end{cases} \tag{8}$$

$$\begin{cases} x_1 \quad - 2x_3 \quad + \; x_5 = 1 \\ \quad\quad\quad\quad x_4 - 6x_5 = 0 \\ \quad x_2 + 5x_3 \quad - 3x_5 = 2 \end{cases} \tag{9}$$

Específicamente, el sistema de ecuaciones (7) no satisface la condición 1 porque el primer coeficiente no nulo de la segunda ecuación es 2; el sistema de ecuaciones (8) no satisface la condición 2 porque x_3, la primera incógnita con coeficiente no nulo de la segunda ecuación, aparece con coeficiente no nulo en la primera ecuación; y el sistema de ecuaciones (9) no satisface la condición 3 porque x_2, la primera incógnita con coeficiente no nulo de la tercera ecuación, no tiene un subíndice mayor que x_4, la primera incógnita con coeficiente no nulo de la segunda ecuación.

Una vez que se ha obtenido un sistema en el que se satisfacen las propiedades 1, 2 y 3, es fácil de resolver para algunas de las incógnitas en términos de las otras (como en el ejemplo anterior). *Sin embargo, si en el curso de la ejecución de las operaciones* 1, 2 *y* 3 *se obtuviera un sistema que tuviera una ecuación de la forma* $0 = c$, *donde* c *no es nula, entonces el sistema original no tiene soluciones.* (Ver el ejemplo 16 a continuación.)

Regresaremos al estudio de sistemas de ecuaciones lineales en el capítulo 3. Al mismo tiempo, expondremos las bases teóricas para este método de solución de sistemas de ecuaciones lineales y su simplificación posterior mediante el uso de las matrices.

Ejemplo 16. Demostraremos que

$$2x^3 - 2x^2 + 12x - 6$$

es una combinación lineal de

$$x^3 - 2x^2 - 5x - 3 \quad \text{y} \quad 3x^3 - 5x^2 - 4x - 9$$

en $P_3(R)$ pero que

$$3x^3 - 2x^2 + 7x + 8$$

no lo es. En el primer caso deseamos encontrar escalares a y b tales que

$$\begin{aligned} 2x^3 - 2x^2 + 12x - 6 &= a(x^3 - 2x^2 - 5x - 3) + b(3x^3 - 5x^2 - 4x - 9) \\ &= (a + 3b)x^3 + (-2a - 5b)x^2 + (-5a - 4b)x + (-3a - 9b). \end{aligned}$$

Esto nos lleva a establecer el siguiente sistema de ecuaciones lineales:

$$\begin{cases} a + 3b = 2 \\ -2a - 5b = -2 \\ -5a - 4b = 12 \\ -3a - 9b = -6. \end{cases}$$

Sumando múltiplos adecuados de la primera ecuación a las otras para eliminar a, encontramos

$$\begin{cases} a + 3b = 2 \\ b = 2 \\ 11b = 22 \\ 0 = 0. \end{cases}$$

Ahora, sumando múltiplos adecuados de la segunda ecuación a las demás se tendrá

$$\begin{cases} a = -4 \\ b = 2 \\ 0 = 0 \\ 0 = 0. \end{cases}$$

Por lo tanto,

$$2x^3 - 2x^2 + 12x - 6 = -4(x^3 - 2x^2 - 5x - 3) + 2(3x^3 - 5x^2 - 4x - 9).$$

En el segundo caso deseamos mostrar que no existen escalares a y b para los cuales

$$3x^3 - 2x^2 + 7x + 8 = a(x^3 - 2x^2 - 5x - 3) + b(3x^3 - 5x^2 - 4x - 9).$$

Como en el caso anterior, obtenemos un sistema de ecuaciones lineales

$$\begin{cases} a + 3b = 3 \\ -2a - 5b = -2 \\ -5a - 4b = 7 \\ -3a - 9b = 8. \end{cases} \qquad (10)$$

Eliminando a, al igual que antes, se tiene

$$\begin{cases} a + 3b = 3 \\ b = 4 \\ 11b = 22 \\ 0 = 17. \end{cases}$$

Pero la presencia de la ecuación inconsistente $0 = 17$ indica que el sistema de ecuaciones (10) no tiene soluciones y, por tanto, $3x^3 - 2x^2 + 7x + 8$ no es una combinación lineal de $x^3 - 2x^2 - 5x - 3$ y $3x^3 - 5x^2 - 4x - 9$.

El conjunto de combinaciones lineales de los elementos de un subconjunto no vacío de un espacio vectorial proporciona otro ejemplo de subespacio, como lo muestra el siguiente resultado.

Teorema 1.7. *Si* S *es un subconjunto no vacío de un espacio vectorial* V, *entonces el conjunto* W, *integrado por todas las combinaciones lineales de elementos de* S, *es un subespacio de* V *más pequeño que contiene a* S *en el sentido de que* W *es un subconjunto de cualquier subespacio de* V *que contiene a* S.

DEMOSTRACIÓN. Primero, emplearemos el Teorema 1.3 para probar que W es un subespacio de V. Como $S \neq \emptyset$, al menos $0 \in$ W. Si y y z son elementos de W, entonces y y z son combinaciones lineales de elementos de S, de manera que existen elementos $x_1, \ldots, x_n$ y $w_1, \ldots, w_m$ en S tales que $y = a_1x_1 + \ldots + a_nx_n$ y $z = b_1w_1 + \ldots + b_mw_m$ para alguna selección de escalares $a_1, \ldots, a_n$ y $b_1, \ldots, b_m$. Ahora bien,

$$y + z = a_1x_1 + \ldots + a_nx_n + b_1w_1 + \ldots + b_mw_m$$

y

$$cy = ca_1x_1 + \ldots + ca_nx_n$$

son combinaciones lineales de elementos de S; luego entonces $y + z$ y cy son elementos de W para cualquier c. Así, tenemos que W es un subespacio de V.

Ahora bien, sea W′ cualquier subespacio de V que contenga a S. Si y es un elemento de W, entonces y es una combinación lineal de elementos de S —digamos $y = a_1x_1 + \ldots + a_nx_n$, donde $a_1, \ldots, a_n \in F$ y $x_1, \ldots, x_n \in S$. Puesto que $S \subseteq$ W′, $x_1, \ldots, x_n \in$ W′. Luego entonces $y = a_1x_1 + \ldots + a_nx_n$ es un elemento de W′, de acuerdo con el ejercicio 24 de la sección 1.3. Como y, un elemento arbitrario de W, pertenece a W′, W $\subseteq$ W′. Esto completa la demostración. ■

Definición. *Al subespacio* W *descrito en el Teorema 1.7 se le llama* subespacio generado por los elementos de S *y se denota por* L(S). *Por conveniencia, definiremos* L(∅) = {*0*}.

Obsérvese que el Teorema 1.7 muestra que x es una combinación lineal de elementos de S si y sólo si x es un elemento de $L(S)$. Luego, por ejemplo, en R^3, $L(\{(1, 0, 0), (0, 1, 0)\})$ es el plano xy.

Definición. *Un conjunto* S *de un espacio vectorial* V genera a V *si* L(S) = V. *En esta situación también podemos decir que* los elementos de S generan a V.

Ejemplo 17. Los vectores (1, 1, 0), (1, 0, 1), y (0, 1, 1) generan a R^3 pues un elemento arbitrario (a_1, a_2, a_3) de R^3 es una combinación lineal de los tres vectores dados; de hecho, los escalares r, s y t para los que

$$r(1, 1, 0) + s(1, 0, 1) + t(0, 1, 1) = (a_1, a_2, a_3)$$

son

$$r = \tfrac{1}{2}(a_1 + a_2 - a_3),\ s = \tfrac{1}{2}(a_1 - a_2 + a_3),\ \text{y}\ t = \tfrac{1}{2}(-a_1 + a_2 + a_3).$$

Ejemplo 18. Los polinomios $x^2 + 3x - 2$, $2x^2 + 5x - 3$ y $-x^2 - 4x + 4$ generan a $P_2(R)$ pues cualquiera de los tres polinomios dados pertenece a $P_2(R)$ y cada polinomio $ax^2 + bx + c$ en $P_2(R)$ es una combinación lineal de los tres; a saber,

$$(-8a + 5b + 3c)(x^2 + 3x - 2) + (4a - 2b - c)(2x^2 + 5x - 3) + (-a + b + c)(-x^2 - 4x + 4) = ax^2 + bx + c.$$

Ejemplo 19. Las matrices

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix},\ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},\ \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},\ \text{y}\ \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

generan a $M_{2\times 2}(R)$ pues un elemento cualquiera

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \text{ de } M_{2\times 2}(R)$$

puede ser expresado como una combinación lineal de las cuatro matrices dadas de la siguiente manera:

$$\begin{aligned} &(\tfrac{1}{3}a_{11} + \tfrac{1}{3}a_{12} + \tfrac{1}{3}a_{21} - \tfrac{2}{3}a_{22})\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \\ &+ (\tfrac{1}{3}a_{11} + \tfrac{1}{3}a_{12} - \tfrac{2}{3}a_{21} + \tfrac{1}{3}a_{22})\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \\ &+ (\tfrac{1}{3}a_{11} - \tfrac{2}{3}a_{12} + \tfrac{1}{3}a_{21} + \tfrac{1}{3}a_{22})\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \\ &+ (-\tfrac{2}{3}a_{11} + \tfrac{1}{3}a_{12} + \tfrac{1}{3}a_{21} + \tfrac{1}{3}a_{22})\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}. \end{aligned}$$

EJERCICIOS

1. Decir si las siguientes expresiones son verdaderas o falsas.
   (a) El vector cero es una combinación lineal de cualquier conjunto no vacío de vectores.
   (b) El subespacio generado por $\varnothing$ es $\varnothing$.
   (c) Si S es un subconjunto de un espacio vectorial V, $L(S)$ es igual a la intersección de todos los subespacios de V que contienen a S.
   (d) Al resolver un sistema de ecuaciones lineales se puede multiplicar una ecuación por una constante.
   (e) Al resolver un sistema de ecuaciones lineales se permite sumar un múltiplo de una ecuación a otra.
   (f) Todo sistema de ecuaciones lineales tiene una solución.

2. Resolver los siguientes sistemas de ecuaciones lineales por el método expuesto en esta sección.

(a) $$\begin{cases} 2x_1 - 2x_2 - 3x_3 = -2 \\ 3x_1 - 3x_2 - 2x_3 + 5x_4 = 7 \\ x_1 - x_2 - 2x_3 - x_4 = -3 \end{cases}$$

(b) $$\begin{cases} 3x_1 - 7x_2 + 4x_3 = 10 \\ x_1 - 2x_2 + x_3 = 3 \\ 2x_1 - x_2 - 2x_3 = 6 \end{cases}$$

(c) $$\begin{cases} x_1 + 2x_2 - x_3 + x_4 = 5 \\ x_1 + 4x_2 - 3x_3 - 3x_4 = 6 \\ 2x_1 + 3x_2 - x_3 + 4x_4 = 8 \end{cases}$$

(d) $$\begin{cases} x_1 + 2x_2 + 2x_3 = 2 \\ x_1 + 8x_3 + 5x_4 = -6 \\ x_1 + x_2 + 5x_3 + 5x_4 = 3 \end{cases}$$

(e) $$\begin{cases} x_1 + 2x_2 - 4x_3 - x_4 + x_5 = 7 \\ -x_1 + 10x_3 - 3x_4 - 4x_5 = -16 \\ 2x_1 + 5x_2 - 5x_3 - 4x_4 - x_5 = 2 \\ 4x_1 + 11x_2 - 7x_3 - 10x_4 - 2x_5 = 7 \end{cases}$$

(f) $$\begin{cases} x_1 + 2x_2 + 6x_3 = -1 \\ 2x_1 + x_2 + x_3 = 8 \\ 3x_1 + x_2 - x_3 = 15 \\ x_1 + 3x_2 + 10x_3 = -5 \end{cases}$$

3. Para cada uno de los siguientes grupos de vectores en R^3, determine si el primer vector puede o no ser expresado como una combinación lineal de los otros dos.

(a) $(-2, 0, 3), (1, 3, 0), (2, 4, -1)$
(b) $(1, 2, -3), (-3, 2, 1), (2, -1, -1)$
(c) $(3, 4, 1), (1, -2, 1), (-2, -1, 1)$
(d) $(2, -1, 0), (1, 2, -3), (1, -3, 2)$
(e) $(5, 1, -5), (1, -2, -3), (-2, 3, -4)$
(f) $(-2, 2, 2), (1, 2, -1), (-3, -3, 3)$

4. Para cada uno de los siguientes grupos de polinomios en $\mathsf{P}_3(R)$, determine si el primer polinomio puede o no ser expresado como una combinación lineal de los otros dos.

(a) $x^3 - 3x + 5,\ x^3 + 2x^2 - x + 1,\ x^3 + 3x^2 - 1$
(b) $4x^3 + 2x^2 - 6,\ x^3 - 2x^2 + 4x + 1,\ 3x^3 - 6x^2 + x + 4$
(c) $-2x^3 - 11x^2 + 3x + 2,\ x^3 - 2x^2 + 3x - 1,\ 2x^3 + x^2 + 3x - 2$
(d) $x^3 + x^2 + 2x + 13,\ 2x^3 - 3x^2 + 4x + 1,\ x^3 - x^2 + 2x + 3$
(e) $x^3 - 8x^2 + 4x,\ x^3 - 2x^2 + 3x - 1,\ x^3 - 2x + 3$
(f) $6x^3 - 3x^2 + x + 2,\ x^3 - x^2 + 2x + 3,\ 2x^3 + x^2 - 3x + 1$

5. En F^n sea e_j el vector cuya coordenada j-ésima es 1 y cuyas otras coordenadas son 0. Demostrar que $\{e_1, e_2, \ldots, e_n\}$ genera a F^n.

6. Mostrar que $\mathsf{P}_n(F)$ puede generarse por $\{1, x, x^2, \ldots, x^n\}$.

7. Mostrar que las matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \text{ y } \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

generan a $\mathsf{M}_{2\times 2}(F)$.

8. Demostrar que si

$$M_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \text{ y } M_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

entonces el subespacio generado por $\{M_1, M_2, M_3\}$ es el conjunto de todas las matrices simétricas de 2×2.

9.* Para cualquier elemento x en un espacio vectorial, demostrar que $L(\{x\}) = \{ax\colon\ a \in F\}$. Interpretar este resultado geométricamente en R^3.

10. Demostrar que un subconjunto W de un espacio vectorial V es un subespacio de V si y sólo si $L(\mathsf{W}) = \mathsf{W}$.

11.* Demostrar que si S_1 y S_2 son subconjuntos de un espacio vectorial V tales que $S_1 \subseteq S_2$, $L(S_1) \subseteq L(S_2)$. En particular, si $S_1 \subseteq S_2$ y $L(S_1) = \mathrm{V}$, se deduce que $L(S_2) = \mathrm{V}$.

12.* Demostrar que si S_1 y S_2 son subconjuntos cualesquiera de un espacio vectorial V, entonces $L(S_1 \cup S_2) = L(S_1) + L(S_2)$.

13. Sean S_1 y S_2 subconjuntos de un espacio vectorial V. Demostrar que $L(S_1 \cap S_2) \subseteq L(S_1) \cap L(S_2)$. Dar un ejemplo en el cual $L(S_1 \cap S_2)$ y $L(S_1) \cap L(S_2)$ sean iguales y un ejemplo donde sean distintas.

1.5 DEPENDENCIA E INDEPENDENCIA LINEAL

Al principio de la sección 1.4, observamos que la ecuación de un plano que pasa por tres puntos no colineales en el espacio, uno de los cuales es el origen, es de la forma $x = t_1u + t_2v$, donde u, $v \in \mathrm{R}^3$ y t_1 y t_2 son escalares. Así, un vector x en R^3 es una combinación lineal de u, $v \in \mathrm{R}^3$ si y sólo si x se ubica en el plano que contiene a u y v. (Ver figura 1.5.) Vemos, por tanto, que en R^3 la amplitud de dos vectores no paralelos tiene una interpretación geométrica sencilla. Se le puede dar una interpretación similar a la amplitud de un vector individual no nulo en R^3. (Ver el ejercicio 9 de la sección 1.4.)

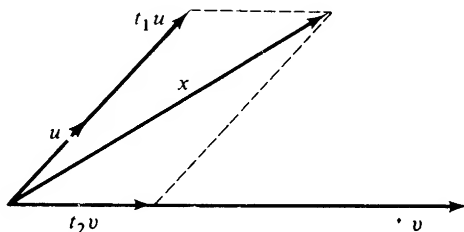


figura 1.5

En la ecuación $x = t_1\, u + t_2v$, x depende de u y v en el sentido de que x es una combinación lineal de u y v. Un conjunto en el que al menos un vector es una combinación lineal de los otros se llama un conjunto linealmente dependiente. Considérese, por ejemplo, el conjunto $S = \{x_1, x_2, x_3, x_4\} \subseteq \mathrm{R}^3$, donde $x_1 = (2, -1, 4)$, $x_2 = (1, -1, 3)$, $x_3 = (1, 1, -1)$, y $x_4 = (1, -2, -1)$. Para determinar si S es linealmente dependiente debemos ver si existe o no un vector en S que sea una combinación lineal de los demás. Ahora bien, el vector x_4 es una combinación lineal de x_1, x_2 y x_3 si y sólo si existen escalares a, b y c tales que

$$x_4 = ax_1 + bx_2 + cx_3,$$

es decir, si y sólo si

$$x_4 = (2a + b + c,\ -a - b + c,\ 4a + 3b - c).$$

Por tanto x_4 es una combinación lineal de x_1, x_2 y x_3 si y sólo si el sistema

$$\begin{cases} 2a + \ b + c = \ \ 1 \\ -a - \ b + c = -2 \\ 4a + 3b - c = -1 \end{cases}$$

tiene solución. El lector deberá verificar que en este caso no existe tal solución. Nótese, sin embargo, que esto no significa que el conjunto S no sea linealmente dependiente, pues es necesario verificar si x_1, x_2 y x_3 pueden o no ser escritos como una combinación de los otros vectores de S. Puede demostrarse, de hecho, que x_3 es una combinación lineal de x_1, x_2 y x_4; específicamente, $x_3 = 2x_1 - 3x_2 + 0x_4$. Así, S es en efecto linealmente dependiente.

Se ve de este ejemplo que la condición para dependencia lineal que se ha dado no es adecuada, porque no todo vector en S necesita ser una combinación lineal de los demás, aun cuando S sea linealmente dependiente. Reformulando la definición de la siguiente manera obtenemos una definición de dependencia más fácil de usar.

Definición. *Un subconjunto* S *de un espacio vectorial* V *es linealmente dependiente si existe un número finito de vectores distintos* $x_1, x_2, \ldots, x_n$ *en* S *y escalares* $a_1, a_2, \ldots, a_n$ *en* F, *no todos cero, tales que* $a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0$. *También se puede describir esta situación diciendo que los elementos de* S *son linealmente dependientes.*

Para demostrar que el subconjunto S de R^3 que hemos definido es linealmente dependiente usando esta definición, debemos encontrar escalares a_1, a_2, a_3 y a_4, no todos nulos, tales que

$$a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 = 0,$$

es decir, tales que

$$(2a_1 + a_2 + a_3 + a_4,\ -a_1 - a_2 + a_3 - 2a_4,\ 4a_1 + 3a_2 - a_3 - a_4) = (0,\ 0,\ 0).$$

Por ello debemos encontrar una solución para el sistema

$$\begin{cases} 2a_1 + \ a_2 + a_3 + \ a_4 = 0 \\ -a_1 - \ a_2 + a_3 - 2a_4 = 0 \\ 4a_1 + 3a_2 - a_3 - \ a_4 = 0 \end{cases}$$

donde no todas las incógnitas valen cero. Como para el caso propuesto antes sabemos que $x_3 = 2x_1 - 3x_2 + 0x_4$, se tiene que $0 = 2x_1 - 3x_2 - x_3 + 0x_4$. De aquí, tenemos que $a_1 = 2$, $a_2 = -3$, $a_3 = -1$ y $a_4 = 0$ es dicha solución.

Por lo tanto se ve que la definición establecida de dependencia lineal requiere de la solución de únicamente un sistema de ecuaciones en vez de dos o más. El lector deberá verificar que las dos condiciones para dependencia lineal que hemos tratado son, de hecho, equivalentes. (Ver ejercicio 10.)

Puede verse fácilmente que, en cualquier espacio vectorial, un subconjunto S que contenga al vector cero debe ser linealmente dependiente. Como $1 \cdot 0 = 0$, el vector cero es una combinación lineal de elementos de S en la que algún coeficiente es no nulo.

Ejemplo 20. En R^4 el conjunto $S = \{(1, 3, -4, 2), (2, 2, -4, 0), (1, -3, 2, -4)\}$ es linealmente dependiente puesto que

$$4(1, 3, -4, 2) - 3(2, 2, -4, 0) + 2(1, -3, 2, -4) = (0, 0, 0, 0).$$

De manera semejante, en $M_{2\times 3}(R)$ el conjunto

$$\left\{\begin{pmatrix} 1 & -3 & 2 \\ -4 & 0 & 5 \end{pmatrix}, \begin{pmatrix} -3 & 7 & 4 \\ 6 & -2 & -7 \end{pmatrix}, \begin{pmatrix} -2 & 3 & 11 \\ -1 & -3 & 2 \end{pmatrix}\right\}$$

es linealmente dependiente puesto que

$$5\begin{pmatrix} 1 & -3 & 2 \\ -4 & 0 & 5 \end{pmatrix} + 3\begin{pmatrix} -3 & 7 & 4 \\ 6 & -2 & -7 \end{pmatrix} - 2\begin{pmatrix} -2 & 3 & 11 \\ -1 & -3 & 2 \end{pmatrix} =$$
$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Definición. *Se dice que un subconjunto S de un espacio vectorial, que no es linealmente dependiente, es linealmente independiente. Como anteriormente, describiremos a menudo esta situación diciendo que los elementos de S son linealmente independientes.*

Nótese que el conjunto vacío es linealmente independiente, puesto que obviamente los conjuntos linealmente dependientes deben ser no vacíos. Más aún, en cualquier espacio vectorial, un conjunto integrado de un solo vector no nulo es linealmente independiente. Si $\{x\}$ es linealmente dependiente, entonces $ax = 0$ para algún escalar a no nulo. Pero entonces

$$x = a^{-1}(ax) = a^{-1}0 = 0.$$

Además, un conjunto S es linealmente independiente si y sólo si las únicas combinaciones lineales de elementos de S iguales a 0 son las com-

binaciones lineales triviales en donde todos los escalares valen cero. Este hecho proporciona un método muy útil para determinar si un conjunto finito es linealmente independiente. Esta técnica se ilustra en el siguiente ejemplo.

Ejemplo 21. Sea x_k el vector en F^n cuyas primeras $k-1$ coordenadas son ceros y cuyas últimas $n-k+1$ coordenadas son 1. Entonces $\{x_1, x_2, \ldots, x_n\}$ es linealmente independiente, porque si $a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0$, igualando las coordenadas correspondientes de la izquierda y derecha de esta igualdad se tiene el siguiente sistema de ecuaciones:

$$\begin{cases} a_1 & = 0 \\ a_1 + a_2 & = 0 \\ a_1 + a_2 + a_3 & = 0 \\ \quad\vdots & \\ a_1 + a_2 + a_3 + \cdots + a_n & = 0. \end{cases}$$

Claramente se ve que la única solución de este sistema es $a_1 = a_2 = \ldots = a_n = 0$.

Los siguientes resultados útiles son consecuencias inmediatas de las definiciones de dependencia e independencia lineal.

Teorema 1.8. *Sea* V *un espacio vectorial y sea* $S_1 \subseteq S_2 \subseteq V$. *Si* S_1 *es linealmente dependiente entonces* S_2 *también lo es.*

DEMOSTRACIÓN. Ejercicio.

Corolario. *Sea* V *un espacio vectorial y sea* $S_1 \subseteq S_2 \subseteq V$. *Si* S_2 *es linealmente independiente entonces* S_1 *también lo es.*

DEMOSTRACIÓN. Ejercicio.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

(a) Si S es un conjunto linealmente dependiente, cada elemento de S es una combinación lineal de otros elementos de S.
(b) Cualquier conjunto que contenga al vector cero es linealmente dependiente.
(c) El conjunto vacío es linealmente dependiente.

(d) Subconjuntos de conjuntos linealmente dependientes son linealmente dependientes.
(e) Subconjuntos de conjuntos linealmente independientes son linealmente independientes.
(f) Si $x_1, x_2, \ldots, x_n$ son linealmente independientes y $a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0$, todos los escalares a_i son iguales a cero.

2. En F^n sea e_j el vector cuya coordenada j-ésima es 1 y las demás son 0. Demostrar que $\{e_1, e_2, \ldots, e_n\}$ es linealmente independiente.

3. Demostrar que el conjunto $\{1, x, x^2, \ldots, x^n\}$ es linealmente independiente en $P_n(F)$.

4. Demostrar que las matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \text{y} \quad \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

son linealmente independientes en $M_{2\times 2}(F)$.

5. Encontrar el conjunto de matrices diagonales linealmente independientes que generan al espacio vectorial de matrices diagonales de 2×2.

6.* Demostrar que $\{x, y\}$ es linealmente dependiente si y sólo si x o y es un múltiplo del otro.

7. Dar un ejemplo de tres vectores linealmente dependientes en R^2 tales que ninguno de los tres es múltiplo de otro.

8. Demostrar el Teorema 1.8 y su corolario.

9. (a) Demostrar que $\{u, v\}$ es linealmente independiente si y sólo si $\{u + v, u - v\}$ es linealmente independiente.
(b) Demostrar que $\{u, v, w\}$ es linealmente independiente si y sólo si $\{u + v, u + w, v + w\}$ es linealmente independiente.

10. Demostrar que un conjunto S es linealmente dependiente si y sólo si $S = \{0\}$ o si existen vectores distintos $y, x_1, x_2, \ldots, x_n$ en S tal que y es una combinación lineal de $x_1, x_2, \ldots, x_n$.

11. Sea $S = \{x_1, x_2, \ldots, x_n\}$ un conjunto finito de vectores. Demostrar que S es linealmente dependiente si y sólo si $x_1 = 0$, o $x_{k+1} \in L(\{x_1, x_2, \ldots, x_k\})$ para alguna $k < n$.

12. Demostrar que un conjunto S de vectores es linealmente independiente si y sólo si cada subconjunto finito de S es linealmente independiente.

13. Sea M una matriz cuadrada triangular superior (como se definió en el ejercicio 12 de la sección 1.3) que tenga términos no nulos en la diagonal. Demostrar que las columnas de M son linealmente independientes.

14. Sean f y g funciones definidas por $f(t) = e^{rt}$ y $g(t) = e^{st}$, donde $r \neq s$. Demostrar que f y g son linealmente independientes en $\mathcal{F}(R, R)$. *Sugerencia:* Suponer que $ae^{rt} + be^{st} = 0$. Hacer $t = 0$ y obtener una ecuación que involucre a y b. Luego diferenciar $ae^{rt} + be^{st} = 0$, y hacer $t = 0$ para tener una segunda ecuación en a y b. Resolver ambas ecuaciones para a y b.

1.6 BASES Y DIMENSION

Un subconjunto S de un espacio vectorial V que sea linealmente independiente y que genere a V posee una propiedad muy útil —cada elemento de V puede ser expresado de una y sólo una manera como combinación lineal de elementos de S. (Esta propiedad será demostrada en el Teorema 1.9.) Es este resultado el que hace que los conjuntos generadores linealmente independientes sean los elementos constructivos de los espacios vectoriales.

Definición. *Una* base β *para un espacio vectorial* V *es un subconjunto linealmente independiente de* V *que genera a* V. *(Si* β *es una base de* V, *diremos a menudo que los elementos de* β *forman una base de* V.)

Ejemplo 22. Recordando que $L(\varnothing) = \{0\}$, se dice que $\varnothing$ es una base para el espacio vectorial $\{0\}$.

Ejemplo 23. En F^n, sea $e_1 = (1, 0, 0, \ldots, 0, 0)$, $e_2 = (0, 1, 0, \ldots, 0, 0)$, $\ldots$, $e_n = (0, 0, 0, \ldots, 0, 1)$; se ve claramente que $\{e_1, e_2, \ldots, e_n\}$ es una base para F^n y se llama *base estándar para* F^n.

Ejemplo 24. En $M_{m\times n}(F)$, sea M^{ij} la matriz cuyo único elemento no nulo es un 1 en el i-ésimo renglón y j-ésima columna. Luego $\{M^{ij}: 1 \leq i \leq m, 1 \leq j \leq n\}$ es una base para $M_{m\times n}(F)$.

Ejemplo 25. En $P_n(F)$ el conjunto $\{1, x, x^2, \ldots, x^n\}$ es una base.

Ejemplo 26. En $P(F)$ el conjunto $\{1, x, x^2, \ldots\}$ es una base.

Observar que el ejemplo 26 muestra que una base no necesariamente debe ser finita. De hecho, veremos más adelante en esta sección que ninguna base para $P(F)$ puede ser finita. Entonces, no todo espacio vectorial tiene una base finita.

El siguiente teorema, que se utilizará frecuentemente en el capítulo 2, muestra la propiedad más importante de una base.

Teorema 1.9. *Sea* V *un espacio vectorial y* $\beta = \{x_1, \ldots, x_n\}$ *un subconjunto de* V. *Luego* β *es una base de* V *si y sólo si cada vector* y *en* V *puede ser expresado de manera única como una combinación lineal de vectores de* β*, es decir, puede ser expresado en la forma*

$$y = a_1x_1 + \ldots + a_nx_n$$

para escalares únicos $a_1, \ldots, a_n$.

DEMOSTRACIÓN. Sea β una base para V. Si $y \in V$, entonces $y \in L(\beta)$ puesto que $L(\beta) = V$. Luego, y es una combinación lineal de los elementos de β. Supóngase que $y = a_1x_1 + \ldots + a_nx_n$ y $y = b_1x_1 + \ldots + b_nx_n$ son dos posibles representaciones de y. Restando la segunda igualdad de la primera se tendrá

$$0 = (a_1 - b_1)x_1 + \ldots + (a_n - b_n)\, x_n.$$

Como β es linealmente independiente, se tiene que $a_1 - b_1 = \ldots = a_n - b_n = 0$. Luego, $a_1 = b_1, \ldots, a_n = b_n$, de tal modo que y sólo puede expresarse como una única combinación lineal de los elementos de β.

La prueba de la proposición recíproca se deja al lector como ejercicio. ■

El Teorema 1.9 muestra que cada vector v en un espacio vectorial V con una base $\beta = \{x_1, \ldots, x_n\}$ puede ser expresado de manera única en la forma $v = a_1x_1 + \ldots + a_nx_n$ para escalares $a_1, \ldots, a_n$ seleccionados adecuadamente. Luego, v determina una única n-dimensional de escalares $(a_1, \ldots, a_n)$ y, recíprocamente, cada n-dimensional de escalares determina un vector único v, al utilizar los términos de la n-dimensional como los coeficientes de una combinación lineal de los vectores de β. Este hecho sugiere que V es similar al espacio vectorial F^n, donde n es el número de vectores de una base para V. En la sección 2.4 veremos que éste es realmente el caso.

Nuestro próximo teorema identificará una gran clase de espacios vectoriales, cada uno de ellos con una base finita. Sin embargo, es necesario que primero probemos un resultado preliminar.

Lema. *Sea* S *un subconjunto linealmente independiente de un espacio vectorial* V, *y sea* x *un elemento de* V *que no está en* S. *Luego,* $S \cup \{x\}$ *es linealmente dependiente si y sólo si* $x \in L(S)$.

DEMOSTRACIÓN. Si $S \cup \{x\}$ es linealmente dependiente, deberán existir vectores $x_1, \ldots, x_n$ en $S \cup \{x\}$ y escalares no nulos $a_1, \ldots, a_n$ tales que $a_1x_1 + \ldots + a_nx_n = 0$. Puesto que S es linealmente independiente, una de las x_i, digamos x_1, es igual a x. Por ello $a_1x + a_2x_2 + \ldots + a_nx_n = 0$,

y así $x = a_1^{-1}(-a_2x_2 - \ldots - a_nx_n)$. Como x es una combinación lineal de $x_2, \ldots, x_n$, que son elementos de S, $x \in L(S)$.

Recíprocamente, supóngase que $x \in L(S)$. Luego, existen vectores $x_1, \ldots, x_n$ en S y escalares $a_1, \ldots, a_n$ tales que $x = a_1x_1 + \ldots + a_nx_n$. Así, $0 = a_1x_1 + \ldots + a_nx_n + (-1)x$, y como $x \neq x_i$ para $i = 1, \ldots, n$, $\{x_1, \ldots, x_n, x\}$ es linealmente dependiente. Por tanto, $S \cup \{x\}$ es linealmente dependiente por el Teorema 1.8. ■

Teorema 1.10. *Si un espacio vectorial* V *es generado por un conjunto finito* S_0, *entonces un subconjunto de* S_0 *es una base para* V. *Y por tanto,* V *tiene una base finita.*

DEMOSTRACIÓN. Si $S_0 = \varnothing$ o $S_0 = \{0\}$, entonces V $= \{0\}$ y $\varnothing$ es un subconjunto de S_0 que es una base para V. De lo contrario, S_0 contendrá un elemento x_1 no nulo. Nótese que $\{x_1\}$ es un conjunto linealmente independiente. Continúese, si es posible, escogiendo elementos $x_2, \ldots, x_r$ en S_0 tales que $\{x_1, x_2, \ldots, x_r\}$ sea linealmente independiente. Como S_0 es un conjunto finito, se debe alcanzar una etapa en la que $S = \{x_1, \ldots, x_r\}$ sea un subconjunto linealmente independiente de S_0 pero que al añadir a S cualquier elemento de S_0 que no esté en S se produzca un conjunto linealmente dependiente. Demostraremos entonces que S es una base para V. Como S es linealmente independiente, basta con demostrar que $L(S) =$ V, pero como $L(S_0) =$ V, de acuerdo con el Teorema 1.7 es suficiente demostrar que $S_0 \subseteq L(S)$. Sea $x \in S_0$. Si $x \in S$, entonces evidentemente $x \in L(S)$. De otra forma, si $x \notin S$, la anterior construcción mostraría que $S \cup \{x\}$ es linealmente dependiente. Así, $x \in L(S)$ de acuerdo con el lema y, por tanto, $S_0 \subseteq L(S)$. ■

El método por el cual se obtuvo la base S en la demostración anterior es una manera útil de obtener bases. Un ejemplo de este procedimiento es el que se da a continuación.

Ejemplo 27. Los elementos $(2, -3, 5)$, $(8, -12, 20)$, $(1, 0, -2)$, $(0, 2, -1)$ y $(7, 2, 0)$ generan a R^3. De entre ellos seleccionaremos una base para R^3. Para empezar, selecciónese cualquier elemento no nulo del conjunto generatriz, digamos $(2, -3, 5)$, como uno de los elementos de la base. Como $4(2, -3, 5) = (8, -12, 20)$, el conjunto $\{(2, -3, 5), (8, -12, 20)\}$ es linealmente dependiente (ejercicio 6, sección 1.5). Por tanto, $(8, -12, 20)$ no será incluido en nuestra base. Como $(1, 0, -2)$ no es múltiplo de $(2, -3, 5)$, y viceversa, el conjunto $\{(2, -3, 5)(1, 0, -2)\}$ es linealmente independiente. Por tanto, $(1, 0, -2)$ puede ser incluido en la base. Procediendo con el siguiente elemento del conjunto generatriz, se deberá excluir o incluir en nuestra base al elemento $(0, 2, -1)$ dependiendo de que el conjunto $\{(2, -3, 5), (1, 0, -2), (0, 2, -1)\}$ sea linealmente dependiente o linealmente independiente. Un cálculo sencillo demuestra que el conjunto es linealmente independiente; luego,

(0, 2, −1) también será incluido en nuestra base. El elemento final del conjunto generatriz (7, 2, 0) será excluido o incluido en nuestra base dependiendo de que {(2, −3, 5), (1, 0, −2), (0, 2, −1), (7, 2, 0)} sea linealmente dependiente o linealmente independiente. Ya que

$$2(2, -3, 5) + 3(1, 0, 0, -2) + 4(0, 2, -1) - (7, 2, 0) = (0, 0, 0),$$

el conjunto es linealmente dependiente y se excluye a (7, 2, 0) de la base. De esta manera, el conjunto {(2, −3, 5), (1, 0, −2), (0, 2, −1)} es una base para R^3.

El siguiente teorema y sus corolarios son quizá los resultados más significativos del capítulo 1.

Teorema 1.11. *Sea* V *un espacio vectorial que tiene una base* β *con exactamente* n *elementos. Sea* $S = \{y_1, \ldots, y_m\}$ *un subconjunto linealmente independiente de* V *que contenga exactamente* m *elementos, donde* $m \leq n$. *Entonces, existe un subconjunto* S_1 *de* β *que contiene exactamente* $n - m$ *elementos tales que* $S \cup S_1$ *genera a* V.

DEMOSTRACIÓN. La demostración se hará por inducción sobre m. Principiaremos la inducción con $m = 0$, pues en este caso $S = \varnothing$, y así $S_1 = \beta$ satisface claramente la conclusión del teorema.

Ahora, supóngase que el teorema es cierto para alguna m tal que $m < n$. Demostraremos que el teorema es cierto para $m + 1$. Sea $S = \{y_1, \ldots, y_m, y_{m+1}\}$ un subconjunto de V linealmente independiente, el cual contiene exactamente $m + 1$ elementos. Como $\{y_1, \ldots, y_m\}$ es linealmente independiente, de acuerdo con el corolario al Teorema 1.8, aplicamos la hipótesis de inducción para concluir que existe un subconjunto $\{x_1, \ldots, x_{n-m}\}$ de β tal que $\{y_1, \ldots, y_m\} \cup \{x_1, \ldots, x_{n-m}\}$ genera a V. Por lo tanto, existirán escalares $a_1, \ldots, a_m, b_1, b_2, \ldots, b_{n-m}$ tales que

$$y_{m+1} = a_1y_1 + \ldots + a_my_m + b_1x_1 + b_2x_2 + \ldots + b_{n-m}x_{n-m}. \quad (11)$$

Obsérvese que algún b_i, tal como b_1, es no nulo, pues de lo contrario la ecuación (11) implicaría que y_{m+1} es una combinación lineal de $y_1, \ldots, y_m$ en contradicción con la suposición de que $\{y_1, \ldots, y_m, y_{m+1}\}$ es linealmente independiente. Resolviendo la ecuación (11) para x_1 se tendrá

$$x_1 = (-b_1^{-1}a_1)y_1 + \ldots + (-b_1^{-1}a_m)y_m - (-b_1^{-1})y_{m+1} + (-b_1^{-1}b_2)x_2 + \ldots + (-b_1^{-1}b_{n-m})x_{n-m}. \quad (12)$$

Entonces $x_1 \in L(\{y_1, \ldots, y_m, y_{m+1}, x_2, \ldots, x_{n-m}\}$ de acuerdo con la ecuación (12), pero como $y_1, \ldots, y_m, x_2, \ldots, x_{n-m}$ son claramente elementos de $L(\{y_1, \ldots, y_m, y_{m+1}, x_2, \ldots, x_{n-m}\})$, se tendrá que

$$\{y_1, \ldots, y_m, x_1, x_2, \ldots, x_{n-m}\} \subseteq L(\{y_1, \ldots, y_m, y_{m+1}, x_2, \ldots, x_{n-m}\}).$$

Por tanto, el Teorema 1.7 implica que

$$L(\{y_1, \ldots, y_m, y_{m+1}, x_2, \ldots, x_{n-m}\}) = \mathsf{V}.$$

Luego, el escoger $S_1 = \{x_2, \ldots, x_{n-m}\}$ demuestra que el teorema es cierto para $m + 1$.

Esto completa la demostración.

Para ilustrar el Teorema 1.11, nótese que $S = \{x^2 - 4, x - 6\}$ es un subconjunto linealmente independiente de $\mathsf{P}_2(F)$. Como $\beta = \{1, x, x^2\}$ es una base de $\mathsf{P}_2(F)$, deberá de existir un subconjunto S_1 de β que contenga $3 - 2 = 1$ elemento tal que $S \cup S_1$ genere a $\mathsf{P}_2(F)$. En este ejemplo cualquier subconjunto de β que contenga un elemento será suficiente para S_1. Con esto se ve que el conjunto S_1 del Teorema 1.11 no necesariamente es único.

Corolario 1. *Sea* V *un espacio vectorial que tiene una base* β *que contenga exactamente* n *elementos. Entonces, cualquier subconjunto linealmente independiente de* V *que contenga exactamente* n *elementos es una base de* V.

DEMOSTRACIÓN. Sea $S = \{y_1, \ldots, y_n\}$ un subconjunto de V linealmente independiente que contiene exactamente n elementos. Aplicando el Teorema 1.11 se ve que existe un subconjunto S_1 de β que contiene $n - n = 0$ elementos tal que $S \cup S_1$ genera a V. Obviamente $S_1 = \varnothing$; luego, S genera a V. Como S es también linealmente independiente, S es una base para V. ■

Ejemplo 28. Los vectores $(1, -3, 2)$, $(4, 1, 0)$ y $(0, 2, -1)$ forman una base para R^3, ya que si

$$a_1(1, -3, 2) + a_2(4, 1, 0) + a_3(0, 2, -1) = (0, 0, 0),$$

entonces a_1, a_2 y a_3 deberán satisfacer el sistema de ecuaciones

$$\begin{cases} a_1 + 4a_2 \phantom{{}+2a_3} = 0 \\ -3a_1 + a_2 + 2a_3 = 0 \\ 2a_1 \phantom{{}+a_2} - a_3 = 0. \end{cases}$$

Pero puede verse fácilmente que la única solución del sistema es $a_1 = 0$, $a_2 = 0$ y $a_3 = 0$. Entonces, $(1, -3, 2)$, $(4, 1, 0)$ y $(0, 2, -1)$ son linealmente independientes y, de acuerdo con el corolario 1, forman una base para R^3.

Corolario 2. *Sea* V *un espacio vectorial que tiene una base* β *con exactamente* n *elementos. Entonces, cualquier subconjunto de* V *que contenga más de* n *elementos es linealmente dependiente. Consecuentemente, cualquier subconjunto de* V *linealmente independiente contiene como máximo* n *elementos.*

DEMOSTRACIÓN. Sea S un subconjunto de V que contiene más de n elementos. Con el fin de llegar a una contradicción supondremos que S es linealmente independiente. Sea S_1 un subconjunto cualquiera de S con exactamente n elementos. Entonces, de acuerdo con el corolario anterior, S_1 es una base de V. Como S_1 es un subconjunto propio de S, podemos tomar un elemento x de S que no sea elemento de S_1. Como S_1 es una base de V, $x \in L(S_1) = \mathsf{V}$. Luego, el lema previo al Teorema 1.10 implica que $S_1 \cup \{x\}$ es linealmente dependiente. Pero $S_1 \cup \{x\} \subseteq S$; luego, S es linealmente dependiente —una contradicción. Se concluye, por tanto, que S es linealmente dependiente. ■

Ejemplo 29. Sea $S = \{x^2 + 7, 8x^2 - 2x, 4x - 3, 7x + 2\}$. Aun cuando se pueda demostrar directamente que S es un subconjunto linealmente dependiente de $\mathsf{P}_2(F)$, esta conclusión se deriva inmediatamente del corolario anterior puesto que $\beta = \{1, x, x^2\}$ es una base para $\mathsf{P}_2(F)$ que contiene menos elementos que S.

Corolario 3. *Sea* V *un espacio vectorial que tiene una base* β *con exactamente* n *elementos. Entonces, toda base para* V *contendrá exactamente* n *elementos.*

DEMOSTRACIÓN. Sea S una base de V. Como S es linealmente independiente tendrá como máximo, de acuerdo con el corolario 2, n elementos. Supóngase que S contiene exactamente m elementos; luego, $m \leq n$. Pero, además, S es una base de V y β es un subconjunto linealmente independiente de V. Entonces, el corolario 2 puede ser aplicado intercambiando los papeles de β y S para dar $n \leq m$. Luego $m = n$. ■

Si un espacio vectorial tiene una base con un número finito de elementos, entonces el corolario anterior establece que el número de elementos en cada base para el espacio es el mismo. Este resultado hace posibles las siguientes definiciones.

Definiciones. *Un espacio vectorial* V *se llama* dimensionalmente finito *si tiene una base que consta de un número finito de elementos; el único número de elementos en cada base de* V *se llama* dimensión de V *y se denota por* dim(V). *Si un espacio vectorial no es dimensionalmente finito, se llama* dimensionalmente infinito.

Los siguientes resultados son consecuencia de los ejemplos 22 a 26.

Ejemplo 30. El espacio vectorial $\{0\}$ tiene dimensión cero.

Ejemplo 31. El espacio vectorial F^n tiene dimensión n.

Ejemplo 32. El espacio vectorial $\mathsf{M}_{m \times n}(F)$ tiene dimensión mn.

Ejemplo 33. El espacio vectorial $\mathsf{P}_n(F)$ tiene una dimensión $n + 1$.

Ejemplo 34. El espacio vectorial $\mathsf{P}(F)$ es dimensionalmente infinito.

Los dos ejemplos siguientes demuestran que la dimensión de un espacio vectorial depende de su campo de escalares.

Ejemplo 35. El espacio vectorial de los números complejos tiene dimensión 1 sobre el campo de los números complejos. (Una base es $\{1\}$.)

Ejemplo 36. El espacio vectorial de los números complejos tiene dimensión 2 sobre el campo de los números reales. (Una base es $\{1, i\}$.)

Corolario 4. *Sea* V *un espacio vectorial de dimensión* n, *y sea* S *un subconjunto de* V *que genera a* V *y contiene como máximo* n *elementos. Entonces,* S *es una base para* V *y, por tanto, contiene exactamente* n *elementos.*

DEMOSTRACIÓN. Existe un subconjunto S_1 de S tal que S_1 es una base de V (Teorema 1.10). Por el corolario 3, S_1 contiene exactamente n elementos. Pero $S_1 \subseteq S$ y S contiene a lo más n elementos, luego $S = S_1$ y S es una base de V. ■

Ejemplo 37. Se tiene del ejemplo 18 y del corolario 4 que $\{x^2 + 3x - 2, 2x^2 + 5x - 3, -x^2 - 4x + 4\}$ es una base para $\mathsf{P}_2(R)$.

Ejemplo 38. Se tiene del ejemplo 19 y del corolario 4 que

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \text{ y } \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

forma una base de $\mathsf{M}_{2\times 2}(R)$.

Corolario 5. *Sea β una base de un espacio vectorial* V *de dimensión* n *y sea* S *un subconjunto linealmente independiente de* V *que contiene* m *elementos. Entonces, existe un subconjunto* S_1 *de β tal que* $S \cup S_1$ *es una base de* V.

DEMOSTRACIÓN. Por el corolario 2 del Teorema 1.11 sabemos que $m \leq n$. Entonces, por el Teorema 1.11, existe un subconjunto S_1 de β que contiene exactamente $n - m$ elementos tal que $S \cup S_1$ genera a V. Es obvio que $S \cup S_1$ contiene a lo más n elementos; así, el corolario 4 implica que $S \cup S_1$ es una base de V. ■

Los Teoremas 1.10 y 1.11, sus cinco corolarios y el ejercicio 11 contienen toda una riqueza de información acerca de las relaciones entre conjuntos linealmente independientes, bases y conjuntos generatrices. Por

esta razón resumiremos los principales resultados de esta sección para situarlos en una mejor perspectiva.

Una base de un espacio vectorial V es un subconjunto linealmente independiente de V que genera a V. Si V tiene una base finita, entonces cualquier base de V contiene el mismo número de vectores. Este número se llama dimensión de V, y se dice que V es dimensionalmente finito. Luego, si la dimensión de V es n, toda base para V contiene exactamente n vectores. Además, cada subconjunto de V linealmente independiente contiene no más de n vectores y puede ser tomado como base de V mediante la inclusión de vectores adecuadamente escogidos. Por otra parte, cada conjunto generatriz de V contiene al menos n vectores y puede ser transformado en una base para V eliminando adecuadamente algunos de los vectores escogidos. La figura 6 describe estas relaciones. Veremos en la sección 2.4 que todo espacio vectorial sobre F de dimensión n es esencialmente el espacio F^n.

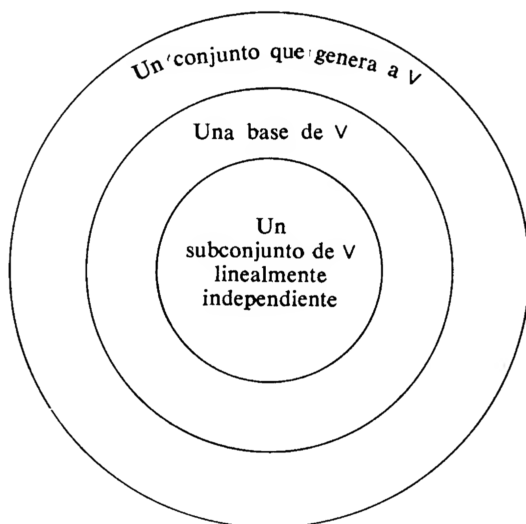


figura 1.6

El siguiente ejemplo ilustra cómo pueden utilizarse estos resultados para obtener una importante conclusión no trivial.

Sean $c_0, c_1, \ldots, c_n$ elementos distintos de un campo infinito F. Los polinomios $f_0(x), f_1(x), \ldots, f_n(x)$, donde

$$f_i(x) = \frac{(x - c_0) \ldots (x - c_{i-1})(x - c_{i+1}) \ldots (x - c_n)}{(c_i - c_0) \ldots (c_i - c_{i-1})(c_i - c_{i+1}) \ldots (c_i - c_n)}$$

$$= \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{x - c_j}{c_i - c_j}$$

se llaman *polinomios de Lagrange (asociados a $c_0, c_1, \ldots, c_n$)*. Tomando a $f_i(x)$ como una función polinomial f_i: $F \to F$, se ve que

$$f_i(c_j) = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j. \end{cases} \tag{13}$$

Se utilizará esta propiedad de los polinomios de Lagrange para demostrar que $\beta = \{f_0, f_1, \ldots, f_n\}$ es un subconjunto linealmente independiente de $\mathsf{P}_n(F)$. Como la dimensión de $\mathsf{P}_n(F)$ es $n + 1$ se tendrá por el corolario 1 del Teorema 1.11 que β es una base de $\mathsf{P}_n(F)$. Para demostrar que β es linealmente independiente, supóngase que:

$$\sum_{i=0}^{n} a_i f_i = 0 \qquad \text{para algunos escalares } a_0, a_1, \ldots, a_n.$$

donde 0 es la función cero. Entonces

$$\sum_{i=0}^{n} a_i f_i(c_j) = 0 \qquad \text{para } j = 0, 1, \ldots, n.$$

pero también

$$\sum_{i=0}^{n} a_i f_i(c_j) = a_j$$

por la ecuación (13). De aquí que $a_j = 0$ para $j = 0, 1, \ldots, n$ y se tiene que β es linealmente independiente.

Como β es una base para $\mathsf{P}_n(F)$, toda función polinomial g en $\mathsf{P}_n(F)$ es una combinación lineal de elementos de β, esto es

$$g = \sum_{i=0}^{n} b_i f_i.$$

Entonces

$$g(c_j) = \sum_{i=0}^{n} b_i f_i(c_j) = b_j;$$

así

$$g = \sum_{i=0}^{n} g(c_i) f_i$$

es la representación única de g como combinación lineal de elementos de β. Esta representación se llama *ecuación de interpolación de Lagrange*. Véase que el argumento anterior muestra que si $b_0, b_1, \ldots, b_n$ son cualesquiera $n + 1$ elementos de F (no necesariamente distintos), entonces la función polinomial

$$g = \sum_{i=0}^{n} b_i f_i$$

es el único elemento de $\mathsf{P}_n(F)$ tal que $g(c_j) = b_j$. Luego entonces, hemos encontrado el único polinomio cuyo grado no excede a n que tiene valores

específicos b_j en puntos dados c_j en su dominio ($j = 0, 1, \ldots, n$). Por ejemplo, construyamos el polinomio real g de grado máximo 2 cuya gráfica contiene a los puntos (1, 8), (2, 5) y (3, −4). (Luego, con la notación anterior, $c_0 = 1$, $c_1 = 2$, $c_2 = 3$, $b_0 = 8$, $b_1 = 5$, y $b_3 = -4$.) Los polinomios de Lagrange asociados a c_0, c_1, y c_2 son

$$f_0(x) = \frac{(x-2)(x-3)}{(1-2)(1-3)} = \frac{1}{2}(x^2 - 5x + 6),$$

$$f_1(x) = \frac{(x-1)(x-3)}{(2-1)(2-3)} = -1(x^2 - 4x + 3),$$

y

$$f_2(x) = \frac{(x-1)(x-2)}{(3-1)(3-2)} = \frac{1}{2}(x^2 - 3x + 2).$$

De aquí, el polinomio deseado es

$$\begin{aligned} g(x) &= \sum_{i=0}^{2} b_i f_i(x) = 8f_0(x) + 5f_1(x) - 4f_2(x) \\ &= 4(x^2 - 5x + 6) - 5(x^2 - 4x + 3) - 2(x^2 - 3x + 2) \\ &= -3x^2 + 6x + 5. \end{aligned}$$

Una consecuencia importante de la ecuación de interpolación de Lagrange es el siguiente resultado: Si $f \in \mathsf{P}_n(F)$ y $f(c_j) = 0$ para $n + 1$ elementos diferentes $c_0, c_1, \ldots, c_n$ en F, f será la función cero.

El siguiente resultado relaciona la dimensión de un subespacio con la dimensión del espacio vectorial que la contiene.

Teorema 1.12. *Sea* W *un subespacio de un espacio vectorial* V *de dimensión* n. *Entonces,* W *es dimensionalmente-finito y dim*(W) $\leq$ n. *Además, si dim*(W) = n, *entonces* W = V.

DEMOSTRACIÓN. Si W = $\{0\}$, entonces W es dimensionalmente finito y dim(W) = $0 \leq n$. De otra manera, existe un elemento no nulo x_1 en W, y así $\{x_1\}$ es un conjunto linealmente independiente. Continuando en esta forma, tómense elementos $x_1, \ldots, x_k$ en W tales que $\{x_1, \ldots, x_k\}$ sea linealmente independiente. Este proceso debe terminar en una etapa donde $\{x_1, \ldots, x_k\}$ sea linealmente independiente pero de manera que al añadir cualquier elemento de W se tenga un conjunto linealmente dependiente (puesto que ningún subconjunto linealmente independiente de V puede contener más de n elementos). Entonces, W tiene una base finita que contiene no más de n elementos; esto es, dim(W) $\leq n$.

Si dim(W) = n, entonces una base para W sería un subconjunto de V linealmente independiente que contuviera n elementos. Pero el corolario 1 del Teorema 1.11 implica que la base para W es también una base para V y se tiene que W = V. ■

Corolario. *Si* W *es un subespacio de un espacio* V *dimensionalmente finito, entonces* W *tiene una base finita y cualquier base para* W *es un subconjunto de una base para* V.

DEMOSTRACIÓN. El teorema muestra que W tiene una base finita S. Si β es alguna base para V, entonces existe un subconjunto S_1 de β tal que $S \cup S_1$ es una base para V (Teorema 1.11). De aquí que S es un subconjunto de una base para V. ■

Podemos utilizar el Teorema 1.12 para analizar geométricamente los subespacios de R^2 y R^3.

Como R^2 tiene dimensión 2 sobre R, los subespacios de R^2 pueden ser solamente de dimensiones 0, 1 ó 2. Los únicos subespacios de dimensiones 0 ó 2 son $\{0\}$ y R^2, respectivamente. Cualquier subespacio de R^2 que tenga dimensión 1 consta de todos los múltiplos escalares de algún vector no nulo en R^2 (ejercicio 9 de la sección 1.4).

Si algún punto de R^2 se identifica de manera natural con un punto del plano Euclidiano, entonces es posible describir los subespacios de R^2 geométricamente: Un subespacio de R^2 de dimensión 0 consta del origen del plano Euclidiano, un subespacio de R^2 de dimensión 1 consta de una recta que pasa por el origen y un subespacio de R^2 que tengan dimensión 2 es todo el plano Euclidiano.

Similarmente, los subespacios de R^3 deben tener dimensión 0, 1, 2 ó 3. Interpretando estas posibilidades geométricamente, vemos que un subespacio de dimensión cero debe ser el origen del sistema coordenado Euclidiano en el espacio, un subespacio de dimensión 1 es una recta que pasa por el origen, un subespacio de dimensión 2 es un plano que pasa por el origen y un subespacio de dimensión 3 es el mismo espacio Euclidiano de 3 dimensiones.

Ejemplo 39. Sea $\mathsf{W} = \{(a_1, \ldots, a_5) \in \mathsf{F}^5\colon a_1 + a_3 + a_5 = 0,\ a_2 = a_4\}$. Entonces W es un subespacio de F^5 con $\{(1, 0, 0, 0, -1), (0, 0, 1, 0, -1), (0, 1, 0, 1, 0)\}$ como una base. Por tanto, la dimensión de W es 3.

Ejemplo 40. El conjunto de las matrices diagonales de $n \times n$ forma un subespacio W de $\mathsf{M}_{n\times n}(F)$. (Ver ejemplo 8.) Una base para W es $\{M^{11}, M^{22}, \ldots, M^{nn}\}$ donde M^{ij} es la matriz definida en el ejemplo 24. Así, la dimensión de W es n.

Ejemplo 41. Vimos en la sección 1.3 que el conjunto de las matrices simétricas de $n \times n$ forma un subespacio W de $\mathsf{M}_{n\times n}(F)$. Una base para W es $\{A^{ij}\colon\ 1 \le i \le j \le n\}$, donde A^{ij} es la matriz de $n \times n$ que tiene 1 en el i-ésimo renglón y la j-ésima columna, 1 en el j-ésimo renglón y i-ésima columna, y 0 en los demás términos. Por tanto, la dimensión de W es $n + (n - 1) + \ldots + 1 = \frac{1}{2}n(n + 1)$.

Ejemplo 42. El conjunto de polinomios de la forma $a_{18}x^{18} + a_{16}x^{16} + \ldots + a_2x^2 + a_0$ donde $a_0, a_2, \ldots, a_{16}, a_{18} \in F$, componen un subespacio W de $\mathsf{P}_{19}(F)$ de dimensión 10 puesto que $\{1, x^2, x^4, \ldots, x^{18}\}$ es una base para W.

Si W_1 y W_2 son subespacios de un espacio vectorial V, vimos en la sección 1.3 que también lo son $\mathsf{W}_1 \cap \mathsf{W}_2$ y $\mathsf{W}_1 + \mathsf{W}_2$. Es natural preguntar si las dimensiones de estos subespacios pueden calcularse directamente a partir de las dimensiones de W_1 y W_2. Desafortunadamente esto no es posible. Existe, sin embargo, una relación entre $\dim(\mathsf{W}_1 + \mathsf{W}_2)$, $\dim(\mathsf{W}_1)$, y $\dim(\mathsf{W}_2)$.

Teorema 1.13. *Sean W_1 y W_2 subespacios dimensionalmente finitos de un espacio vectorial* V. *Entonces,* $\mathsf{W}_1 + \mathsf{W}_2$ *es dimensionalmente finito y*

$$\dim(\mathsf{W}_1 + \mathsf{W}_2) = \dim(\mathsf{W}_1) + \dim(\mathsf{W}_2) - \dim(\mathsf{W}_1 \cap \mathsf{W}_2).$$

DEMOSTRACIÓN. Como $\mathsf{W}_1 \cap \mathsf{W}_2$ es un subespacio de un espacio dimensionalmente finito W_1, $\mathsf{W}_1 \cap \mathsf{W}_2$ tiene una base finita $\beta_0 = \{x_1, \ldots, x_k\}$ (Teorema 1.12). Usemos el corolario del Teorema 1.12 para encontrar $\beta_1 = \{y_1, \ldots, y_r\}$ y $\beta_2 = \{z_1, \ldots, z_m\}$ tales que $\beta_0 \cup \beta_1$ sea una base para W_1 y $\beta_0 \cup \beta_2$ sea una base para W_2. Demostraremos que $\beta_0 \cup \beta_1 \cup \beta_2 = \{x_1, \ldots, x_k, y_1, \ldots, y_r, z_1, \ldots, z_m\}$ es una base para $\mathsf{W}_1 + \mathsf{W}_2$. Se seguirá que $\mathsf{W}_1 + \mathsf{W}_2$ es dimensionalmente finita y que

$$\begin{aligned}\dim(\mathsf{W}_1 + \mathsf{W}_2) &= k + r + m = (k + r) + (k + m) - k\\ &= \dim(\mathsf{W}_1) + \dim(\mathsf{W}_2) - \dim(\mathsf{W}_1 \cap \mathsf{W}_2).\end{aligned}$$

Para demostrar que $\beta_0 \cup \beta_1 \cup \beta_2$ es una base para $\mathsf{W}_1 + \mathsf{W}_2$, demostraremos primero que $\beta_0 \cup \beta_1 \cup \beta_2$ es linealmente independiente. Supóngase que

$$a_1x_1 + \ldots + a_kx_k + b_1y_1 + \ldots + b_ry_r + c_1z_1 + \ldots + c_mz_m = 0$$

para algunos escalares $a_1, \ldots, a_k, b_1, \ldots, b_r, c_1, \ldots, c_m$. Sea

$$v_0 = a_1x_1 + \ldots + a_kx_k, \qquad v_1 = b_1y_1 + \ldots + b_ry_r,$$

y

$$v_2 = c_1z_1 + \ldots + c_mz_m;$$

obsérvese que $v_0 \in \mathsf{W}_1 \cap \mathsf{W}_2$, $v_1 \in \mathsf{W}_1$, y $v_2 \in \mathsf{W}_2$. La igualdad anterior puede expresarse como $v_0 + v_1 + v_2 = 0$; así, $v_0 + v_1 = -v_2$. En esta última igualdad el miembro izquierdo es un elemento de W_1 y el miembro derecho es un elemento de W_2. Entonces $-v_2$ es tanto un elemento de W_1 como de W_2, esto es, $-v_2 \in \mathsf{W}_1 \cap \mathsf{W}_2$. Como $\{x_1, \ldots, x_k\}$ es una base para $\mathsf{W}_1 \cap \mathsf{W}_2$, existen escalares $d_1, \ldots, d_k$ tales que $-v_2 = d_1x_1 + \ldots + d_kx_k$. Ahora bien,

$$\begin{aligned} 0 &= v_0 + v_1 + v_2 \\ &= (a_1x_1 + \ldots + a_kx_k) + (b_1y_1 + \ldots + b_ry_r) \\ &\quad + (-d_1x_1 - \ldots - d_kx_k) \\ &= (a_1 - d_1)x_1 + \ldots + (a_k - d_k)x_k + b_1y_1 + \ldots + b_ry_r. \end{aligned}$$

Así tenemos una combinación lineal de elementos de $\beta_0 \cup \beta_1$ que es igual al vector cero; pero $\beta_0 \cup \beta_1$ es un conjunto linealmente independiente, y así $a_1 - d_1 = \ldots = a_k - d_k = b_1 = \ldots = b_r = 0$. De aquí que $v_1 = 0$. Entonces

$$0 = v_0 + v_1 + v_2 = v_0 + v_2 = a_1x_1 + \ldots + a_kx_k + c_1z_1 + \ldots + c_mz_m,$$

de manera que una combinación lineal de elementos de $\beta_0 \cup \beta_2$ es igual al vector cero. Como antes, el hecho de que $\beta_0 \cup \beta_2$ sea un conjunto linealmente independiente implica que $a_1 = \ldots = a_k = c_1 = \ldots = c_m = 0$. Como $a_1 = \ldots = a_k = b_1 = \ldots = b_r = c_1 = \ldots = c_m = 0$, hemos demostrado que $\beta_0 \cup \beta_1 \cup \beta_2$ es linealmente independiente.

Falta demostrar que $\beta_0 \cup \beta_1 \cup \beta_2$ genera a $W_1 + W_2$. Pero ahora tenemos que $L(\beta_0 \cup \beta_1) = W_1$ y $L(\beta_0 \cup \beta_2) = W_2$ puesto que $\beta_0 \cup \beta_1$ y $\beta_0 \cup \beta_2$ son bases para W_1 y W_2, respectivamente. Pero

$$\begin{aligned} L(\beta_0 \cup \beta_1 \cup \beta_2) &= L((\beta_0 \cup \beta_1) \cup (\beta_0 \cup \beta_2)) \\ &= L(\beta_0 \cup \beta_1) + L(\beta_0 \cup \beta_2) \\ &= W_1 + W_2 \end{aligned}$$

por el ejercicio 12 de la sección 1.4. De aquí que $\beta_0 \cup \beta_1 \cup \beta_2$ genera a $W_1 + W_2$, lo cual completa la demostración. ■

Como una consecuencia inmediata de este resultado, se tiene el siguiente corolario de utilidad.

Corolario. *Sean W_1 y W_2 subespacios dimensionalmente finitos de un espacio vectorial V tales que $V = W_1 + W_2$. Luego, V es la suma directa de W_1 y W_2 si y sólo si*

$$dim(V) = dim(W_1) + dim(W_2).$$

Ejemplo 43. Sea c un elemento de un campo infinito F, sea W_1 el conjunto de todas las funciones constantes en $P_n(F)$, y defínase como $W_2 = \{f(x) \in P_n(F);\ f(c) = 0\}$. Puede verse fácilmente que W_1 y W_2 son subespacios de $P_n(F)$ y que $P_n(F) = W_1 \oplus W_2$. (Obsérvese que para cualesquiera $f(x) \in P_n(F)$, $g(x) = f(c) \in W_1$, $h(x) = f(x) - f(c) \in W_2$, y $f(x) = g(x) + h(x)$.) Como la función constante $p(x) = 1$ claramente constituye una base para W_1, se deduce del corolario anterior que

$$\dim(W_2) = \dim(P_n(F)) - \dim(W_1) = (n + 1) - 1 = n.$$

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

(a) El espacio vectorial cero no tiene base.
(b) Todo espacio vectorial generado por un conjunto finito tiene una base.
(c) Todo espacio vectorial tiene una base finita.
(d) Un espacio vectorial no puede tener más de una base.
(e) Si un espacio vectorial tiene una base finita, entonces el número de vectores en todas las bases es el mismo.
(f) La dimensión de $\mathsf{P}_n(F)$ es n.
(g) La dimensión de $\mathsf{M}_{m\times n}(F)$ es $m + n$.
(h) Suponer que V es un espacio vectorial dimensionalmente finito, que S_1 es un subconjunto linealmente independiente de V y que S_2 es un subconjunto de V que genera a V. Luego, S_1 no puede tener más elementos que S_2.
(i) Si S genera al espacio vectorial V, entonces todo vector en V puede escribirse como una combinación lineal de elementos de S de una sola manera.
(j) Todo subespacio de un espacio dimensionalmente finito es dimensionalmente finito.
(k) Si V es un espacio vectorial de dimensión n, entonces V tiene exactamente un subespacio de dimensión 0 y exactamente un subespacio de dimensión n.
(l) Si W_1 y W_2 son subespacios dimensionalmente finitos de un espacio vectorial, entonces $\dim(\mathsf{W}_1 + \mathsf{W}_2) = \dim(\mathsf{W}_1) + \dim(\mathsf{W}_2)$.

2. Determinar cuáles de los siguientes conjuntos son bases para R^3.

(a) $\{(1, 0, -1), (2, 5, 1), (0, -4, 3)\}$
(b) $\{(2, -4, 1), (0, 3, -1), (6, 0, -1)\}$
(c) $\{(1, 2, -1), (1, 0, 2), (2, 1, 1)\}$
(d) $\{(-1, 3, 1), (2, -4, -3), (-3, 8, 2)\}$
(e) $\{(1, -3, -2), (-3, 1, 3), (-2, -10, -2)\}$

3. Determinar cuáles de los siguientes conjuntos son bases para $\mathsf{P}_2(R)$.

(a) $\{-1 - x + 2x^2, 2 + x - 2x^2, 1 - 2x + 4x^2\}$
(b) $\{1 + 2x + x^2, 3 + x^2, x + x^2\}$
(c) $\{1 + 4x - 2x^2, -2 + 3x - x^2, -3 - 12x + 6x^2\}$
(d) $\{-1 + 2x + 4x^2, 3 - 4x - 10x^2, -2 - 5x - 6x^2\}$
(e) $\{1 + 2x - x^2, 4 - 2x + x^2, -1 + 18x - 9x^2\}$

4. ¿Generan los polinomios $x^3 - 2x^2 + 1$, $4x^2 - x + 3$ y $3x - 2$ a $\mathsf{P}_3(R)$? Justifique su respuesta.

5. ¿Es $\{(1, 4, -6), (1, 5, 8), (2, 1, 1), (0, 1, 0)\}$ un subconjunto linealmente independiente de R^3? Justifique su respuesta.

6. Dar tres bases diferentes para F^2 y para $\mathsf{M}_{2\times 2}(F)$.

7. Los vectores $x_1 = (2, -3, 1)$, $x_2 = (1, 4, -2)$, $x_3 = (-8, 12, -4)$, $x_4 = (1, 37, -17)$, y $x_5 = (-3, -5, 8)$ generan a R^3. Encontrar un subconjunto de $\{x_1, x_2, x_3, x_4, x_5\}$ que sea una base para R^3.

8. Sea V el espacio vectorial que consta de todos los vectores de R^5 para los cuales la suma de las coordenadas es cero. Los vectores

$$\begin{aligned} x_1 &= (2, -3, 4, -5, 2), & x_2 &= (-6, 9, -12, 15, -6), \\ x_3 &= (3, -2, 7, -9, 1), & x_4 &= (2, -8, 2, -2, 6), \\ x_5 &= (-1, 1, 2, 1, -3), & x_6 &= (0, -3, -18, 9, 12), \\ x_7 &= (1, 0, -2, 3, -2), & x_8 &= (2, -1, 1, -9, 7) \end{aligned}$$

generan a V. Encontrar un subconjunto de $\{x_1, \ldots, x_8\}$ que sea una base para V.

9. Los vectores $x_1 = (1, 1, 1, 1)$, $x_2 = (0, 1, 1, 1)$, $x_3 = (0, 0, 1, 1)$, y $x_4 = (0, 0, 0, 1)$ forman una base para F^4. Encontrar la única representación de un vector arbitrario (a_1, a_2, a_3, a_4) en F^4 como combinación lineal de los vectores x_1, x_2, x_3, y x_4.

10. Sea

$$\mathsf{V} = \mathsf{M}_{2\times 2}(F), \qquad \mathsf{W}_1 = \left\{\begin{pmatrix} a & b \\ c & a \end{pmatrix} \in \mathsf{V}\colon \quad a, b, c \in F\right\}$$

y

$$\mathsf{W}_2 = \left\{\begin{pmatrix} 0 & a \\ -a & b \end{pmatrix} \in \mathsf{V}\colon \quad a, b \in F\right\}.$$

Demostrar que W_1 y W_2 son subespacios de V y encontrar las dimensiones de W_1, W_2, $\mathsf{W}_1 + \mathsf{W}_2$, y $\mathsf{W}_1 \cap \mathsf{W}_2$.

11.* Sea V un espacio vectorial de dimensión n y sea S un subconjunto de V que genera a V.

(a) Demostrar que S contiene al menos n elementos.
(b) Demostrar que un subconjunto de S es una base para V. (Tenga cuidado de no suponer que S sea finito.)

12. Sean W_1 y W_2 subespacios de un espacio vectorial V de dimensiones m y n, respectivamente, donde $m \geq n$. Demostrar que $\dim(\mathsf{W}_1 \cap \mathsf{W}_2 \leq n$ y

$\dim(W_1 + W_2) \leq m + n$. Dar ejemplos de subespacios de R^3 donde cada desigualdad se convierta en igualdad.

13. Sea $\{x, y\}$ una base de un espacio vectorial V. Mostrar que tanto $\{x + y, x - y\}$ como $\{ax, by\}$ son bases para V, donde a y b son escalares arbitrarios no nulos.

14. Suponer que V es un espacio vectorial con una base $\{x_1, x_2, x_3\}$. Demostrar que $\{x_1 + x_2 + x_3, x_2 + x_3, x_3\}$ también es una base para V.

15. El conjunto de soluciones para el sistema

$$\begin{cases} x_1 - 2x_2 + x_3 = 0 \\ 2x_1 - 3x_2 + x_3 = 0 \end{cases}$$

es un subespacio de R^3. Encontrar una base para este subespacio.

16. Encontrar bases para los siguientes subespacios de F^5:

$$W_1 = \{(a_1, a_2, a_3, a_4, a_5) \in F^5: \quad a_1 - a_3 - a_4 = 0\}$$

y

$$W_2 = \{(a_1, a_2, a_3, a_4, a_5) \in F^5: \quad a_2 = a_3 = a_4,\ a_1 + a_5 = 0\}.$$

¿Cuáles son las dimensiones de W_1 y W_2?

17. El conjunto de todas las matrices de $n \times n$ cuya traza es igual a cero es un subespacio W de $M_{n \times n}(F)$. (Ver Ejemplo 1.1.) Encontrar una base para W. ¿Cuál es la dimensión de W?

18. El conjunto de todas las matrices triangulares de $n \times n$ es un subespacio W de $M_{n \times n}(F)$. (Ver Ejercicio 12 de la Sección 1.3.) Encontrar una base para W. ¿Cuál es la dimensión de W?

19. El conjunto de todas las matrices antisimétricas de $n \times n$ es un subespacio W de $M_{n \times n}(F)$. (Ver Ejercicio 25 de la Sección 1.3.) Encontrar una base para W. ¿Cuál es la dimensión de W?

20. (a) Sean W_1 y W_2 subespacios de un espacio vectorial V tales que $V = W_1 \oplus W_2$. Si β_1 y β_2 son bases para W_1 y W_2, respectivamente, demostrar que $\beta_1 \cap \beta_2 = \varnothing$ y que $\beta_1 \cup \beta_2$ es una base para V.

(b) Recíprocamente, sean β_1 y β_2 bases disjuntas para subespacios W_1 y W_2, respectivamente, de un espacio vectorial V. Demostrar que si $\beta_1 \cup \beta_2$ es una base para V, entonces $V = W_1 \oplus W_2$.

21. Completar la demostración del Teorema 1.9.

22. Sea W un subespacio de un espacio vectorial dimensionalmente finito V. Determinar la dimensión del espacio vectorial V/W, el espacio cociente de V módulo W. (Ver Ejercicio 29 de la Sección 1.3.) Justifique su respuesta.

23. Encontrar una base para el espacio vectorial de sucesiones no nulas en un campo F. (Ver ejemplo 5.)

24. Demostrar que si W_1 es un subespacio cualquiera de un espacio vectorial dimensionalmente finito V, entonces existe un subespacio W_2 de V: tal que $V = W_1 \oplus W_2$.

25. Demostrar que un espacio vectorial es dimensionalmente infinito si y sólo si contiene un subconjunto infinito linealmente independiente.

1.7 SUBCONJUNTOS MAXIMOS LINEALMENTE INDEPENDIENTES*

En esta sección extenderemos algunos resultados importantes de la sección 1.6 de manera que incluyan espacios vectoriales dimensionalmente infinitos. Nuestra meta principal es demostrar que todo espacio vectorial tiene una base. Este resultado es fundamental para el estudio de espacios vectoriales dimensionalmente infinitos, ya que a menudo es extremadamente difícil construir explícitamente una base para tales espacios.

La dificultad que surge al expandir los teoremas de la sección anterior a espacios dimensionalmente infinitos es que el *principio de inducción matemática,* que jugó un papel fundamental en muchas de las demostraciones de la sección 1.6, ya no es válido. En vez de ello, utilizaremos un principio más general llamado *principio de maximidad,* el que requiere de la siguiente terminología.

Definición. *Sea $\mathcal{F}$ una familia de conjuntos. Un miembro* M *de $\mathcal{F}$ se llama* máximo (en relación con la inclusión de conjunto), *si ningún miembro de $\mathcal{F}$ contiene propiamente a* M.

Ejemplo 44. Sea $\mathcal{F}$ la familia de todos los subconjuntos de un conjunto no vacío S ($\mathcal{F}$ se denomina *conjunto potencia de S*). Se puede ver fácilmente que S es el elemento máximo de $\mathcal{F}$.

Definición. *Una colección de conjuntos $\mathcal{C}$ se denomina* cadena *si, para cada par de conjuntos* A *y* B *en $\mathcal{C}$, se tiene que* $A \subseteq B$ *o* $B \subseteq A$.

Ejemplo 45. Sea A_n el conjunto que consta de los enteros 1, 2, . . . , n. Entonces $\mathcal{C} = \{A_n\colon \ n = 1, 2, 3, \ldots\}$ es una cadena; de hecho $A_m \subseteq A_n$ si y sólo si $m \leq n$.

Con esta terminología ya podemos expresar el principio de maximidad.

Principio de Maximidad. *Sea $\mathcal{F}$ una familia de conjuntos. Si, para cada cadena $\mathcal{C} \subseteq \mathcal{F}$, existe un miembro de $\mathcal{F}$ que contiene a cada uno de los miembros de $\mathcal{C}$, entonces $\mathcal{F}$ contiene un elemento máximo.*

Como el principio de maximidad garantiza la existencia de elementos máximos en una familia de conjuntos, será útil reformular la definición de una base en términos de la propiedad de maximidad. Demostraremos posteriormente que esta reformulación es equivalente a la definición original de base.

Definición. *Sea* S *un subconjunto de un espacio vectorial* V. *Un* subconjunto máximo linealmente independiente *de* S *es un subconjunto* B *de* S *que satisface las siguientes condiciones:*

(a) B *es linealmente independiente.*
(b) *Cualquier subconjunto de* S *que contenga propiamente a* B *es linealmente dependiente.*

Ejemplo 46. El ejemplo 16 muestra que $\{x^3 - 2x^2 - 5x - 3,\ 3x^3 - 5x^2 - 4x - 9\}$ es un subconjunto máximo linealmente independiente de

$$S = \{2x^3 - 2x^2 + 12x - 6,\ x^3 - 2x^2 - 5x - 3,\ 3x^3 - 5x^2 - 4x - 9\}$$

en $P_3(R)$. En este caso, sin embargo, se puede demostrar fácilmente que cualquier subconjunto de dos elementos de S es un subconjunto máximo linealmente independiente de S. De aquí que los subconjuntos máximos linealmente independientes de un conjunto no necesariamente son únicos.

Una base β para un espacio vectorial V es un subconjunto máximo linealmente independiente de V, ya que:

(a) β es, por definición, linealmente independiente.
(b) Si $x \in$ V, $x \notin \beta$, entonces $\beta \cup \{x\}$ es linealmente independiente de acuerdo con el lema del Teorema 1.10, puesto que $L(\beta) =$ V.

Nuestro siguiente resultado muestra que la recíproca de este argumento también es verdadera.

Teorema 1.14. *Sea* S *un subconjunto de un espacio vectorial* V *tal que* S *genera a* V, *y sea β un subconjunto máximo linealmente independiente de* S. *Entonces β es una base para* V.

DEMOSTRACIÓN. Puesto que β es linealmente independiente, es suficiente demostrar que β genera a V. Supóngase que $S \subseteq L(\beta)$; entonces existe $x \in S$ tal que $x \notin L(\beta)$. Pero entonces el lema del Teorema 1.10 implica

que $\beta \cup \{x\}$ es linealmente independiente, lo que es una contradicción a la maximidad de β. Luego entonces $S \subseteq L(\beta)$. Por tanto, como $L(S) = \mathsf{V}$, se tiene del ejercicio 11 de la sección 1.4 que $L(\beta) = \mathsf{V}$. ■

Corolario. *Un subconjunto β de un espacio vectorial* V *es una base para* V *si y sólo si β es un subconjunto máximo linealmente independiente de* V.

En vista de nuestro corolario anterior, podemos llegar a asegurar que todo espacio vectorial tiene una base al demostrar que todo espacio vectorial contiene un subconjunto máximo linealmente independiente. Este resultado se deduce de una manera inmediata a partir de nuestro siguiente teorema.

Teorema 1.15. *Sea* S *un subconjunto linealmente independiente de un espacio vectorial* V. *Existe un subconjunto máximo linealmente independiente de* V *que contiene a* S.

DEMOSTRACIÓN. Sea $\mathcal{F}$ la familia de todos los subconjuntos linealmente independientes de V que contienen a S. Utilizaremos el principio de maximidad para demostrar que $\mathcal{F}$ contiene un elemento máximo. Con el objeto de aplicar el principio de maximidad debemos demostrar que si $\mathcal{C}$ es una cadena en $\mathcal{F}$, entonces existe un miembro U de $\mathcal{F}$ que contiene a cada miembro de $\mathcal{C}$. Demostraremos que U, la unión de los miembros de $\mathcal{C}$, es el conjunto deseado. Como es evidente que U contiene a cada miembro de $\mathcal{C}$, basta con demostrar que $U \in \mathcal{F}$, es decir, que U es un subconjunto linealmente independiente de V que contiene a S. Ahora bien, cada elemento de $\mathcal{C}$ es un subconjunto de V que contiene a S; de aquí $S \subseteq U \subseteq \mathsf{V}$. Para demostrar que U es linealmente independiente, sean $u_1, \ldots, u_n$ vectores en U y $c_1, \ldots, c_n$ escalares tales que $c_1u_1 + \ldots + c_nu_n = 0$. Como $u_i \in U$ para $i = 1, \ldots, n$, existen conjuntos A_i en $\mathcal{C}$ tales que $u_i \in A_i$. Pero como $\mathcal{C}$ es una cadena, uno de los conjuntos $A_1, \ldots, A_n$, por ejemplo A_k, contiene a los demás. Entonces $u_1, \ldots, u_n \in A_k$ para $i = 1, \ldots, n$. Sin embargo, A_k es un conjunto linealmente independiente, de manera que $c_1u_1 + \ldots + c_nu_n = 0$ implica que $c_1 = \ldots = c_n = 0$. Por lo tanto, U es linealmente independiente.

El principio de maximidad implica que $\mathcal{F}$ contiene un elemento máximo, y se ve fácilmente que este elemento máximo es un subconjunto máximo linealmente independiente de V que contiene a S. ■

Corolario. *Todo espacio vectorial tiene una base.*

Puede demostrarse, de una manera semejante a la del Corolario 3 del Teorema 1.11, que toda base para un espacio vectorial dimensionalmente infinito tiene la misma cardinalidad. (Consultar, por ejemplo, a N. Jacobson, *Lecturas sobre Algebra Lineal,* III, pág. 154, D. Van Nostrand Company, Nueva York, 1964.)

Los ejercicios 2 a 5 extienden otros resultados de la sección 1.6 para incluir espacios dimensionalmente infinitos.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

   (a) Toda familia de conjuntos contiene un elemento máximo.
   (b) Toda cadena contiene un elemento máximo.
   (c) Si una familia de conjuntos tiene un elemento máximo, entonces tal elemento máximo es único.
   (d) Si una cadena de conjuntos tiene un elemento máximo, entonces tal elemento máximo es único.
   (e) Una base de un espacio vectorial es un subconjunto máximo linealmente independiente de ese espacio vectorial.
   (f) Un subconjunto máximo linealmente independiente de un espacio vectorial es una base para tal espacio vectorial.

2. Sea W un subespacio de un espacio vectorial V (no necesariamente dimensionalmente finito). Demostrar que cualquier base para W es un subconjunto de una base para V.

3. Demostrar la siguiente versión dimensionalmente infinita del Teorema 1.9: Sea β un subconjunto de un espacio vectorial V dimensionalmente infinito. Entonces β es una base para V si y sólo si para cada vector y no nulo en V existen vectores únicos $x_1, \ldots, x_n$ en β y escalares no nulos únicos $c_1, \ldots, c_n$ tales que $y = c_1x_1 + \ldots + c_nx_n$.

4. Demostrar la siguiente generalización del Teorema 1.10: Sean S_1 y S_2 subconjuntos de un espacio vectorial V tales que $S_1 \subseteq S_2$. Si S_1 es linealmente independiente y S_2 genera a V, entonces existe una base β para V tal que $S_1 \subseteq \beta \subseteq S_2$. *Sugerencia:* Aplicar el principio de maximidad a la familia de todos los subconjuntos linealmente independientes de S_2 que contienen a S_1 y proceder como se hizo en la demostración del Teorema 1.15.

5. Demostrar la siguiente generalización del Teorema 1.11. Sea β una base para un espacio vectorial V y sea S un subconjunto linealmente independiente de V. Existe un subconjunto S_1 de β tal que $S \cup S_1$ es una base de V.

INDICE DE DEFINICIONES PARA EL CAPITULO 1

Capítulo 2

Transformaciones lineales y matrices

En el capítulo 1 desarrollamos la teoría de espacios vectoriales abstractos con bastante detalle. Ahora es natural considerar a aquellas funciones definidas en espacios vectoriales que en cierto sentido "conservan" la estructura; estas funciones especiales se denominan "transformaciones lineales" y son abundantes en las matemáticas puras como en las aplicadas. En el cálculo, las operaciones de diferenciación e integración nos proporcionan dos de los ejemplos más importantes de transformaciones lineales (ver ejemplos 1 y 2). Estos dos ejemplos nos permiten reformular muchos de los problemas de ecuaciones diferenciales e integrales en términos de transformaciones lineales en espacios vectoriales particulares (ver las secciones 2.7 y 5.2).

En geometría, las rotaciones, reflexiones y proyecciones (ver ejemplos 5, 6 y 7) nos proporcionan otra clase de transformaciones lineales, las que utilizaremos posteriormente para estudiar los movimientos rígidos en R^n (sección 7.8).

En los capítulos restantes veremos ejemplos adicionales de transformaciones lineales en las ciencias físicas y sociales.

A lo largo de este capítulo supondremos que todos los espacios vectoriales están definidos sobre un campo ordinario F.

2.1 TRANSFORMACIONES LINEALES, ESPACIOS NULOS Y RANGOS

En esta sección consideraremos un gran número de ejemplos de transformaciones lineales, muchas de las cuales serán estudiadas con más detalle en secciones posteriores.

Definición. *Sean* V *y* W *espacios vectoriales (sobre* F*). Una función* $\mathsf{T}\colon \mathsf{V} \to \mathsf{W}$ *se llama* transformación lineal de V en W *si para toda* $x, y \in \mathsf{V}$ *y* $c \in F$ *tenemos que*

(a) $T(x + y) = T(x) + T(y)$.
(b) $T(cx) = cT(x)$.

A menudo denominaremos a T simplemente *lineal*. El lector deberá verificar los siguientes hechos sobre la función $T: V \rightarrow W$.

1. Si T es lineal, entonces $T(0) = 0$.
2. T es lineal si y sólo si $T(ax + y) = aT(x) + T(y)$ para toda $x, y \in V$ y $a \in F$.
3. T es lineal si y sólo si para $x_1, \ldots, x_n \in V$ y $a_1, \ldots, a_n \in F$ tenemos que

$$T(\sum_{i=1}^{n} a_i x_i) = \sum_{i=1}^{n} a_i T(x_i).$$

Generalmente utilizaremos la propiedad 2 para demostrar que una transformación dada es lineal.

Ejemplo 1. Sea $V = P_n(R)$ y $W = P_{n-1}(R)$ y defínase $T: V \rightarrow W$ mediante $T(f) = f'$, donde f' es la derivada de f. Para demostrar que T es lineal, sean g y h vectores en $P_n(R)$ y $a \in R$. Tenemos que $T(ag + h) =$ $= (ag + h)' = ag' + h' = aT(g) + T(h)$. Entonces, de acuerdo con la propiedad 2, T es lineal.

Ejemplo 2. Sea $V = C(R)$ el espacio vectorial de funciones continuas de variable real en R. Sean $a, b \in R$, $a < b$ y defínase $T: V \rightarrow R$ mediante $T(f) = \int_a^b f(t)dt$ para toda $f \in V$. Entonces, por las propiedades elementales de las integrales, T es una transformación lineal.

Dos ejemplos muy importantes de transformaciones lineales que pueden aparecer a menudo en el resto del libro y que, por tanto, merecen tener una notación propia, son la transformación identidad y la transformación cero.

Para espacios vectoriales V y W (sobre F) definimos la *transformación identidad* $I: V \rightarrow V$ mediante $I_V(x) = x$ para toda $x \in V$ y la *transformación cero* $T_0: V \rightarrow W$ por $T_0(x) = 0$ para toda $x \in V$. Es evidente que ambas transformaciones son lineales. A menudo escribiremos I en vez de I_V.

Veremos ahora algunos ejemplos adicionales de transformaciones lineales.

Ejemplo 3. Defínase

$$T: R^2 \rightarrow R^2 \text{ por } T(a_1, a_2) = (2a_1 + a_2, a_1).$$

Para demostrar que T es lineal, sean

$$x, y \in R^2, \quad x = (b_1, b_2), \quad y = (d_1, d_2), \quad \text{y sea} \quad c \in F.$$

Como

$$cx + y = (cb_1 + d_1,\ cb_2 + d_2),$$

tenemos

$$\mathsf{T}(cx + y) = (2(cb_1 + d_1) + cb_2 + d_2,\ cb_1 + d_1).$$

También

$$\begin{aligned} c\mathsf{T}(x) + \mathsf{T}(y) &= c(2b_1 + b_2,\ b_1) + (2d_1 + d_2,\ d_1) \\ &= (2cb_1 + cb_2 + 2d_1 + d_2,\ cb_1 + d_1) \\ &= (2(cb_1 + d_1) + cb_2 + d_2,\ cb_1 + d_1). \end{aligned}$$

Por lo tanto, T es lineal.

Ejemplo 4. Defínase T: $\mathsf{M}_{m\times n}(F) \to \mathsf{M}_{m\times n}(F)$ mediante $\mathsf{T}(A) = A^t$, donde A^t es tal como se definió en la Sección 1.3. Entonces, T es una transformación lineal por el ejercicio 3 de la Sección 1.3.

Como veremos en las Secciones 7.7 y 7.8, las aplicaciones del álgebra lineal a la geometría son vastas y variadas. La razón principal de esto es que la mayor parte de las transformaciones geométricas son lineales. Tres transformaciones particulares que ahora consideraremos son la rotación, la reflexión y la proyección. Dejaremos al lector las demostraciones de linealidad.

Ejemplo 5. Para $0 \leq \theta < 2\pi$ definamos T_θ: $\mathsf{R}^2 \to \mathsf{R}^2$ mediante

$$\mathsf{T}_\theta(a_1, a_2) = (a_1 \cos\theta - a_2 \operatorname{sen}\theta,\ a_1 \operatorname{sen}\theta + a_2 \cos\theta).$$

T_θ se denomina *rotación en* θ. (Ver Fig. 2.1(a).)

Ejemplo 6. Defínase T: $\mathsf{R}^2 \to \mathsf{R}^2$ mediante $\mathsf{T}(a_1, a_2) = (a_1, -a_2)$. T se denomina *reflexión en torno al eje* x. (Ver Fig. 2.1(b).)

Ejemplo 7. Defínase T: $\mathsf{R}^2 \to \mathsf{R}^2$ mediante $\mathsf{T}(a_1, a_2) = (a_1, 0)$. T se denomina *proyección sobre el eje* x. (Ver Fig. 2.1(c).) Nótese que si hacemos $\mathsf{W}_1 = \{(a, 0):\ a \in R\}$ y $\mathsf{W}_2 = \{(0, a):\ a \in R\}$ entonces $\mathsf{R}^2 = \mathsf{W}_1 \oplus \mathsf{W}_2$,

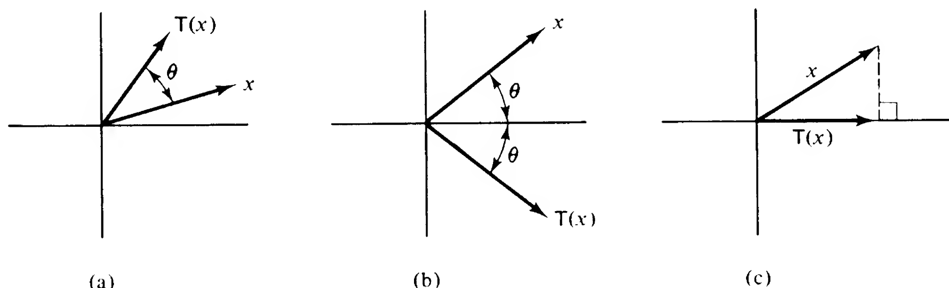


figura 2.1

por lo que para toda $x \in R^2$ existen vectores únicos $x_1 \in W_1$ y $x_2 \in W_2$ tales que $x = x_1 + x_2$ y $T(x) = x_1$.

El ejemplo 7 sugiere la siguiente definición.

Definición. *Sean* V *un espacio vectorial y* W_1 *un subespacio de* V. *Una función* $T: V \to V$ *se llama* proyección sobre W_1 *si*

(a) *Existe un subespacio* W_2 *tal que* $V = W_1 \oplus W_2$

(b) *Para* $x = x_1 + x_2$, *donde* $x_1 \in W_1$ *y* $x_2 \in W_2$, *tenemos* $T(x) = x_1$.

Se deja como ejercicio para el lector demostrar que T es lineal y que $W_1 = \{x: T(x) = x\}$.

Ahora supóngase que existe un subespacio $W'_2 \neq W_2$ tal que $V = W_1 \oplus W'_2$. Defínase $U: V \to V$ mediante $U(x) = x_1$ donde $x = x_1 + x'_2$, $x_1 \in W_1$, y $x_2 \in W'_2$. Entonces, U es otra proyección sobre W_1 y de nuevo $W_1 = \{x: U(x) = x\}$. Por ejemplo, en el Ejemplo 7 sea

$$W'_2 = \{(a, a): a \in R\},$$

tal que

$$(a_1, a_2) = (a_1 - a_2, 0) + (a_2, a_2) \quad \text{y} \quad U(a_1, a_2) = (a_1 - a_2, 0).$$

Entonces existen tantas proyecciones en W_1 como subespacios en W_2 que satisfacen $V = W_1 \oplus W_2$. Veremos en el Capítulo 7 que la proyección descrita en la figura 2.1(c) es la proyección "natural" a ser estudiada. Este tipo de proyección se llamará "proyección ortogonal" y está determinada de manera única por el subespacio W_1.

En el Ejercicio 14 de la Sección 2.3 se dará una caracterización de las proyecciones, la que nos permitirá determinar fácilmente si una transformación lineal es o no una proyección.

Ahora pondremos atención a dos conjuntos muy importantes asociados con las transformaciones lineales: el "rango" y el "espacio nulo". La determinación de estos conjuntos nos permitirá examinar más de cerca las propiedades intrínsecas de una transformación lineal.

Definiciones. *Sean* V *y* W *espacios vectoriales y sea* $T: V \to W$ *lineal. Definimos al* espacio nulo *(o* kernel) $N(T)$ *de* T *como el conjunto de todos los vectores* x *en* V *tal que* $T(x) = 0$; *es decir,* $N(T) = \{x \in V: T(x) = 0\}$. *Definimos al rango (o* imagen) $R(T)$ *como el subconjunto de* W *que consta de todas las imágenes (bajo* T*) de los elementos de* V; *es decir,* $R(T) = \{T(x): x \in V\}$.

Ejemplo 8. Sean V y W espacios vectoriales y sean $I: V \to V$ y $T_0: V \to W$ respectivamente las transformaciones identidad y cero, tal como se definieron anteriormente. Entonces $N(I) = \{0\}$, $R(I) = V$, $N(T_0) = V$ y $R(T_0) = \{0\}$.

Ejemplo 9. Sea $T: R^3 \to R^2$ definida mediante $T(a_1, a_2, a_3) = (a_1 - a_2, 2a_3)$. Se deja como ejercicio verificar que $N(T) = \{(a, a, 0): a \in R\}$ y $R(T) = R^2$.

En el ejemplo 7 se ve fácilmente que los subespacios $W_1 = R(T)$ y $W_2 = N(T)$. El teorema siguiente nos dice que éste es el caso para todas las proyecciones.

Teorema 2.1. *Sea V un espacio vectorial y W_1 un subespacio de V. Sea T una proyección sobre W_1, y sea W_2 tal como en la def.nición de proyección. Entonces*

$$W_1 = R(T) \quad y \quad W_2 = N(T).$$

DEMOSTRACIÓN. Como se observó anteriormente, $W_1 = \{x: T(x) = x\}$. Por lo tanto, $W_1 \subseteq R(T)$. Si $x \in R(T)$, entonces: $x = T(y)$ para alguna $y \in V$. Pero $y = y_1 + y_2$, donde $y_1 \in W_1$ y $y_2 \in W_2$, y entonces $x = y_1$. Por lo tanto, $W_1 = R(T)$.

Como es evidente que $W_2 \subseteq N(T)$, únicamente necesitamos demostrar que $N(T) \subseteq W_2$. Con este fin, sea $x \in N(T)$. Entonces $x = x_1 + x_2$ con $x_1 \in W_1$ y $x_2 \in W_2$. Así, $0 = T(x) = x_1$, y por lo tanto $x = x_2 \in W_2$. ■

Este teorema nos dice que W_2 queda determinada de manera única por la proyección de T sobre W_1. Además, como T es una proyección sobre su rango, utilizaremos sencillamente el término "proyección" sin mencionar al subespacio W_1.

Acabamos de observar en el caso en que T es una proyección que N(T) y R(T) son subespacios de V. Puede obtenerse el mismo resultado con cualquier transformación lineal.

Teorema 2.2. *Sean V y W espacios vectoriales y $T: V \to W$ lineal. Entonces N(T) y R(T) son subespacios de V y W, respectivamente.*

DEMOSTRACIÓN. Para aclarar la notación, usaremos los símbolos 0_V y 0_W para denominar, respectivamente, a los vectores cero de V y W.

Como $T(0_V) = 0_W$, tenemos que $0_V \in N(T)$. Sean $x, \in N(T)$ y $c \in F$. Entonces $T(x + y) = T(x) + T(y) = 0_W + 0_W = 0_W$, y $T(cx) = cT(x) = c0_W = 0_W$. Por lo tanto $x + y \in N(T)$ y $cx \in N(T)$, de manera que N(T) es un subespacio de V.

Como $T(0_V) = 0_W$, tenemos que $0_W \in R(T)$. Ahora sean $x, y \in R(T)$ y $c \in F$. Entonces, existen v y w en V tales que $T(v) = x$ y $T(w) = y$. Así, $T(v + w) = T(v) + T(w) = x + y$, y $T(cv) = cT(v) = cx$. Por lo tanto, $x + y \in R(T)$ y $cx \in R(T)$, de manera que R(T) es un subespacio de W. ■

Tal como en el Capítulo 1, mediremos el "tamaño" de un subespacio por su dimensión. Los dos subespacios anteriores son tan importantes que dedicaremos atención especial a sus dimensiones respectivas.

Definiciones. *Sean* V *y* W *espacios vectoriales y sea* $T: V \to W$ *lineal. Si* N(T) *y* R(T) *son dimensionalmente finitos, entonces definimos la* nulidad de T, *expresada como* nulidad(T), *y el* rango de T, *expresado como* rango(T), *como las dimensiones de* N(T) *y* R(T), *respectivamente.*

Reflexionando sobre la acción de una transformación lineal, vemos intuitivamente que mientras más grande es la nulidad menor es el rango. En otras palabras, mientras más vectores van a dar al cero, menor es el rango. El mismo razonamiento nos dirá que mientras más grande sea el rango, menor es la nulidad. Este balance entre el rango y la nulidad se hará más preciso en el teorema siguiente.

Teorema 2.3. *Sean* V *y* W *espacios vectoriales y sea* $T: V \to W$ *lineal. Si* V *es dimensionalmente finito, entonces* $nulidad(T) + rango(T) = dim(V)$.

DEMOSTRACIÓN. Supóngase que $\dim(V) = n$, y sea $\{x_1, \ldots, x_k\}$ una base para N(T). Por el corolario del Teorema 1.12 podemos extender $\{x_1, \ldots, x_k\}$ a una base $\beta = \{x_1, \ldots, x_n\}$ para V. Demostraremos que el conjunto $S = \{T(x_{k+1}), \ldots, T(x_n)\}$ es una base para R(T).

Primero demostraremos que S genera a R(T). Sea $y \in R(T)$. Entonces existe $x \in V$ tal que $y = T(x)$. Como β es una base para V, tenemos que

$$x = \sum_{i=1}^{n} a_i x_i \qquad \text{para algunas } a_1, \ldots, a_n \in F.$$

Como T es lineal se tiene que

$$y = T(x) = \sum_{i=1}^{n} a_i T(x_i) = \sum_{i=k+1}^{n} a_i T(x_i) \in L(S)$$

La última igualdad se obtiene de que $x_1, \ldots, x_k \in N(T)$.

Ahora demostraremos que S es linealmente independiente. Supóngase que

$$\sum_{i=k+1}^{n} b_i T(x_i) = 0 \qquad \text{para } b_{k+1}, \ldots, b_n \in F.$$

De nuevo, utilizando el hecho de que T es lineal, tenemos que

$$T\left(\sum_{i=k+1}^{n} b_i x_i\right) = 0.$$

Entonces

$$\sum_{i=k+1}^{n} b_i x_i \in N(T).$$

Por lo tanto existen $c_1, \ldots, c_k \in F$ tales que

$$\sum_{i=k+1}^{n} b_i x_i = \sum_{i=1}^{k} c_i x_i \quad \text{o bien} \quad \sum_{i=1}^{k} (-c_i) x_i + \sum_{i=k+1}^{n} b_i x_i = 0.$$

Como β es una base para V, tenemos que $b_i = 0$ para toda i. Por lo tanto, S es linealmente independiente. ■

La demostración anterior permite obtener el corolario siguiente.

Corolario. *Sean* V *y* W *espacios vectoriales y sea* $\mathsf{T}\colon \mathsf{V} \to \mathsf{W}$ *lineal. Si* V *tiene una base* $\beta = \{x_1, \ldots, x_n\}$, *entonces* $\mathsf{R}(\mathsf{T}) = \mathsf{L}(\mathsf{T}(\beta)) = \mathsf{L}(\{\mathsf{T}(x_1), \ldots, \mathsf{T}(x_n)\})$.

Este corolario nos dice que la imagen de una base para el dominio de una transformación lineal es un conjunto generador para el rango de la transformación. Por lo tanto, este corolario proporciona un método para encontrar una base para el rango de una transformación lineal. Emplearemos esta técnica en el ejemplo siguiente.

Ejemplo 10. Defínase la transformación lineal $\mathsf{T}\colon \mathsf{P}_2(R) \to \mathsf{M}_{2\times 2}(R)$ mediante

$$\mathsf{T}(f) = \begin{pmatrix} f(1) - f(2) & 0 \\ 0 & f(0) \end{pmatrix}.$$

Como $\beta = \{1, x, x^2\}$ es una base para $\mathsf{P}_2(R)$, tenemos

$$\begin{aligned} \mathsf{R}(\mathsf{T}) &= L\,(\mathsf{T}(\beta)) = L(\{\mathsf{T}(1), \mathsf{T}(x), \mathsf{T}(x^2)\}) \\ &= L\left(\left\{\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} -3 & 0 \\ 0 & 0 \end{pmatrix}\right\}\right) \\ &= L\left(\left\{\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}\right\}\right). \end{aligned}$$

Entonces, hemos encontrado una base para $\mathsf{R}(\mathsf{T})$ y vemos que $\text{rango}(\mathsf{T}) = 2$. En virtud del Teorema 2.3 tenemos que $\text{nulidad}(\mathsf{T}) + 2 = 3$, y entonces $\text{nulidad}(\mathsf{T}) = 1$.

El lector debería repasar los conceptos de "uno-a-uno" y "sobreyectividad" los cuales se encuentran en el Apéndice B, pues, interesantemente, para una transformación lineal ambos conceptos están íntimamente ligados con el rango y la nulidad de la transformación. Esto quedará demostrado en los dos teoremas siguientes.

Teorema 2.4. *Sean* V *y* W *espacios vectoriales y sea* $\mathsf{T}\colon \mathsf{V} \to \mathsf{W}$ *lineal. Entonces* T *es uno-a-uno si y sólo si* $\mathsf{N}(\mathsf{T}) = \{0\}$.

DEMOSTRACIÓN. Supóngase que T es uno-a-uno y que $x \in \mathsf{N}(\mathsf{T})$. Entonces $\mathsf{T}(x) = 0 = \mathsf{T}(0)$. Como T es uno-a-uno, tenemos que $x = 0$ y por tanto $\mathsf{N}(\mathsf{T}) = \{0\}$.

Ahora supóngase que $N(T) = \{0\}$ y que $T(x) = T(y)$. Entonces $0 = T(x) - T(y) = T(x - y)$. Por lo tanto, $x - y \in N(T) = \{0\}$ y entonces $x - y = 0$, o sea $x = y$. Esto significa que T es uno-a-uno. ■

El lector debe observar que el Teorema 2.4 nos permite concluir que la transformación definida en el Ejemplo 10 no es uno-a-uno.

Sorpresivamente, las condiciones para que una transformación sea uno-a-uno y sobreyectiva son equivalentes en un caso especial de importancia.

Teorema 2.5. *Sean* V *y* W *espacios vectoriales de dimensiones iguales (finitas) y sea* $T: V \to W$ *lineal. Entonces* T *es uno-a-uno si y sólo si* T *es sobreyectiva.*

DEMOSTRACIÓN. Del Teorema 2.3 tenemos que

$$\text{nulidad}(T) + \text{rango}(T) = \dim(V).$$

Ahora, mediante el uso del Teorema 2.4, tenemos que T es uno-a-uno si y sólo si $N(T) = \{0\}$, si y sólo si $\text{nulidad}(T) = 0$, si y sólo si $\text{rango}(T) = \dim(V)$, si y sólo si $\text{rango}(T) = \dim(W)$, si y sólo si $\dim(R(T)) = \dim(W)$. En virtud del Teorema 1.12 esta igualdad es equivalente a $R(T) = W$—la definición de T si ésta es sobreyectiva. ■

La linealidad de T en los Teoremas 2.4 y 2.5 es esencial puesto que es fácil construir ejemplos de funciones de R en R que no sean uno-a-uno pero que sean sobreyectivas, y viceversa.

Los siguientes dos ejemplos hacen uso de los teoremas anteriores para ver si una transformación lineal dada es uno-a-uno o sobreyectiva.

Ejemplo 11. Defínase

$$T: P_2(R) \to P_3(R) \quad \text{mediante} \quad T(f)(x) = 2f'(x) + \int_0^x 3f(t)\,dt.$$

Ahora

$$R(T) = L(\{T(1), T(x), T(x^2)\}) = L(\{3x,\ 2 + \tfrac{3}{2}x^2,\ 4x + x^3\}).$$

Por lo que $\text{rango}(T) = 3$. Como $\dim(P_3(R)) = 4$, T no es sobreyectiva. Del Teorema 2.3, $\text{nulidad}(T) + 3 = 3$; por tanto, $\text{nulidad}(T) = 0$ y entonces $N(T) = \{0\}$. Luego, de acuerdo con el Teorema 2.4, T es uno-a-uno.

Ejemplo 12. Defínase

$$T: F^2 \to F^2 \quad \text{mediante} \quad T(a_1, a_2) = (a_2 + a_1, a_1).$$

Es fácil comprobar que $N(T) = \{0\}$; entonces T es uno-a-uno, por lo que el Teorema 2.5 nos dice que T debe ser sobreyectiva.

Nuestro siguiente teorema proporciona una caracterización de las transformaciones lineales uno-a-uno como aquellas transformaciones que conservan la independencia lineal.

Teorema 2.6. *Sean* V *y* W *espacios vectoriales, y sea* T: $V \to W$ *lineal. Entonces* T *es uno-a-uno si y sólo si* T *lleva subconjuntos linealmente independientes de* V *a subconjuntos linealmente independientes de* W.

DEMOSTRACIÓN. Ejercicio.

Corolario. *Sean* V *y* W *espacios vectoriales y sea* T: $V \to W$ *una transformación lineal uno-a-uno. Supóngase que* S *es un subconjunto de* V. *Entonces* S *es linealmente independiente si y sólo si* T(S) *es linealmente independiente*

DEMOSTRACIÓN. Ejercicio.

Ejemplo 13. Defínase

$$T: P_2(R) \to R^3 \quad \text{mediante} \quad T(a_0 + a_1x + a_2x^2) = (a_0, a_1, a_2).$$

Se ve claramente que T es uno-a-uno. Sea $S = \{2 - x + 3x^2, x + x^2, 1 - 2x^2\}$. Entonces S es linealmente independiente en $P_2(R)$ si y sólo si

$$T(S) = \{2, -1, 3), (0, 1, 1), (1, 0, -2)\}$$

es linealmente independiente en R^3.

En el Ejemplo 13 transferimos un problema del espacio vectorial de los polinomios a un problema en el espacio vectorial de las ternas (3-dimensionales). Esta técnica será explotada después más ampliamente.

Una de las propiedades más importantes de las transformaciones lineales es que quedan completamente definidas por su acción sobre una base. Este resultado, que se obtiene del siguiente teorema y su corolario, se utilizará frecuentemente a lo largo del libro.

Teorema 2.7. *Sean* V *y* W *espacios vectoriales y supóngase que* V *es un espacio vectorial dimensionalmente finito con una base* $\{x_1, \ldots, x_n\}$. *Para cualquier subconjunto* $\{y_1, \ldots, y_n\}$ *de* W *existe exactamente una transformación lineal* T: $V \to W$ *tal que* $T(x_i) = y_i$ *para* $i = 1, \ldots, n$.

DEMOSTRACIÓN. Sea $x \in V$. Entonces

$$x = \sum_{i=1}^{n} a_i x_i,$$

donde $a_1, \ldots, a_n$ son escalares únicos. Defínase

$$T: V \to W \quad \text{mediante} \quad T(x) = \sum_{i=1}^{n} a_i y_i.$$

(a) T es lineal, pues supóngase que u, $v \in \mathsf{V}$ y $d \in F$. Entonces podemos escribir

$$u = \sum_{i=1}^{n} b_i x_i \quad \text{y} \quad v = \sum_{i=1}^{n} c_i x_i.$$

Ahora bien,

$$du + v = \sum_{i=1}^{n} (db_i + c_i)x_i.$$

Entonces

$$\mathsf{T}(du + v) = \sum_{i=1}^{n} (db_i + c_i)y_i = d\sum_{i=1}^{n} b_i y_i + \sum_{i=1}^{n} c_i y_i = d\mathsf{T}(u) + \mathsf{T}(v).$$

(b) Es evidente que

$$\mathsf{T}(x_i) = y_i \qquad \text{para } i = 1, \ldots, n.$$

(c) T es única, porque supóngase que $\mathsf{U}\colon \mathsf{V} \to \mathsf{W}$ es lineal y $\mathsf{U}(x_i) = y_i$ para $i = 1, \ldots, n$. Entonces para $x \in \mathsf{V}$ con

$$x = \sum_{i=1}^{n} a_i x_i$$

tenemos

$$\mathsf{U}(x) = \sum_{i=1}^{n} a_i \mathsf{U}(x_i) = \sum_{i=1}^{n} a_i y_i = \mathsf{T}(x).$$

Por lo tanto U = T. ■

Corolario. *Sean* V *y* W *espacios vectoriales y supóngase que* V *es dimensionalmente finito con una base* $\{x_1, \ldots, x_n\}$. *Si* U, T: $\mathsf{V} \to \mathsf{W}$ *son lineales y* $\mathsf{U}(x_i) = \mathsf{T}(x_i)$ *para* $i = 1, \ldots, n$, *entonces* U = T.

Ejemplo 14. Defínase T: $\mathsf{R}^2 \to \mathsf{R}^2$ mediante $\mathsf{T}(a_1, a_2) = (2a_2 - a_1, 3a_1)$, y supóngase que U: $\mathsf{R}^2 \to \mathsf{R}^2$ es lineal. Si sabemos que $\mathsf{U}(1, 2) = (3, 3)$ y $\mathsf{U}(1, 1) = (1, 3)$, entonces U = T. Esto se deduce del corolario y del hecho de que $\{(1, 2), (1, 1)\}$ es una base para R^2.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas. De aquí en adelante, V y W son espacios vectoriales dimensionalmente finitos (sobre F) y T es una función de V en W.

   (a) Si T es lineal, entonces T conserva las sumas y productos por escalares.
   (b) Si $\mathsf{T}(x + y) = \mathsf{T}(x) + \mathsf{T}(y)$ entonces T es lineal.

(c) T es uno- a-uno si y sólo si $N(T) = \{0\}$.
(d) Todas las proyecciones deben ser lineales.
(e) Si T es lineal, entonces $T(0_V) = 0_W$.
(f) Si T es lineal, entonces nulidad(T) + rango(T) = dim(W).
(g) Si T es lineal, entonces lleva subconjuntos linealmente independientes de V a subconjuntos linealmente independientes de W.
(h) Si T, U: $V \to W$ son lineales y concuerdan en una base de V, entonces T = U.
(i) Dados $x_1, x_2 \in V$ y $y_1, y_2 \in W$, existe una transformación lineal T: $V \to W$ tal que $T(x_1) = y_1$ y $T(x_2) = y_2$.

Para los ejercicios 2 a 6, demostrar que T es una transformación lineal y encontrar bases para N(T) y R(T). Luego, calcular la nulidad y el rango de T y verificar el Teorema 2.3. Finalmente, emplear los teoremas adecuados de esta sección para determinar si T es uno-a-uno o sobreyectiva.

2. T: $R^3 \to R^2$; $T(a_1, a_2, a_3) = (a_1 - a_2, 2a_3)$.

3. T: $R^2 \to R^3$; $T(a_1, a_2) = (a_1 + a_2, 0, 2a_1 - a_2)$.

4. T: $M_{2\times 3}(F) \to M_{2\times 2}(F)$; $T\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} = \begin{pmatrix} 2a_{11} - a_{12} & a_{13} + 2a_{12} \\ 0 & 0 \end{pmatrix}$.

5. T: $P_2(R) \to P_3(R)$; $T(f(x)) = xf(x) + f'(x)$.

6. T: $M_{n\times n} \to F$; $T(A) = \text{tr}(A)$. Recuérdese que

$$\text{tr}(A) = \sum_{i=1}^{n} A_{ii}.$$

7. Verificar los enunciados 1, 2 y 3 de la página 64.

8. Verificar que las transformaciones definidas en los Ejemplos 5, 6 y 7 son lineales.

9. Para las siguientes T: $R^2 \to R^2$, decir por qué T *no* es lineal.

(a) $T(a_1, a_2) = (1, a_2)$

(b) $T(a_1, a_2) = (a_1, a_1)$
(c) $T(a_1, a_2) = (\text{sen}\, a_1, 0)$
(d) $T(a_1, a_2) = (|a_1|, a_2)$
(e) $T(a_1, a_2) = (a_1 + 1, a_2)$

10. Supóngase que T: $R^2 \to R^2$ es lineal y que $T(1, 0) = (1, 4)$ y $T(1, 1) = (2, 5)$. ¿Qué es $T(2, 3)$? ¿T es uno-a-uno?

11. Demostrar que existe una transformación lineal T: $R^2 \to R^3$ tal que $T(1, 1) = (1, 0, 2)$ y $T(2, 3) = (1, -1, 4)$. ¿Qué es $T(8, 11)$?

12. ¿Existe una transformación lineal T: $R^3 \to R^2$ tal que $T(1, 0, 3) = (1, 1)$ y $T(-2, 0, -6) = (2, 1)$?

13. Demostrar el Teorema 2.6 y su corolario.

14. Supóngase que T es una proyección en un subespacio W de un espacio vectorial V. Demostrar que $W = \{x \in V\colon\ T(x) = x\}$.

15. Recordando la definición de $P(R)$ dada en la Sección 1.2, defínase

$$T\colon\ P(R) \to P(R) \quad \text{y} \quad T(f)(x) = \int_0^x f(t)\,dt.$$

Demostrar que T es uno-a-uno pero no sobreyectiva.

16. Sean V y W espacios vectoriales dimensionalmente finitos y T: $V \to W$ lineal.

(a) Demostrar que si $\dim(V) < \dim(W)$, entonces T no puede ser sobreyectiva.
(b) Demostrar que si $\dim(V) > \dim(W)$, entonces T no puede ser uno-a-uno.

17. Dar un ejemplo de una transformación lineal T: $R^2 \to R^2$ tal que $N(T) = R(T)$.

18. Dar un ejemplo de transformaciones lineales diferentes T y U tales que $N(T) = N(U)$ y $R(T) = R(U)$.

19. Sean V y W espacios vectoriales con subespacios V_1 y W_1, respectivamente. Si T: $V \to W$ es lineal, demostrar que $T(V_1)$ es un subespacio de W y $\{x \in V\colon\ T(x) \in W_1\}$ es un subespacio de V.

20. Sea W un subespacio de un espacio vectorial dimensionalmente finito V. Demostrar que existe una proyección sobre W.

21. Sean V y W espacios vectoriales y sea T: $V \to W$ lineal. Sea $\{y_1, \ldots, y_k\}$ un subconjunto linealmente independiente de $R(T)$. Si $S = \{x_1, \ldots, x_k\}$ se selecciona de tal manera que $T(x_i) = y_i$ para $i = 1, \ldots, k$, demostrar que S es linealmente independiente.

22. Sea T: $R^3 \to R$ lineal. Demostrar que existen escalares a, b y c tales que $T(x, y, z) = ax + by + cz$ para toda $(x, y, z) \in R^2$. ¿Se puede generalizar este resultado para T: $F^n \to F$? Enunciar y demostrar un resultado semejante para T: $F^n \to F^m$.

23. Sea T: $R^3 \to R$ lineal. Describir geométricamente las posibilidades para el espacio nulo de T. *Sugerencia:* Usar el Ejercicio 22.

24. Sea V un espacio vectorial y sea $T: V \to V$ lineal. Se dice que un subespacio W de V es T-*invariante* si $T(x) \in W$ para toda $x \in W$, es decir, $T(W) \subseteq W$.

(a) Demostrar que los subespacios $\{0\}$, V, R(T) y N(T) son todos T-invariantes.

(b) Si W es un subespacio T-invariante de V, defínase a $T_W: W \to W$ mediante $T_W(x) = T(x)$ para toda $x \in W$. Demostrar que T_W es lineal.

(c) Si T es una proyección sobre W, demostrar que W es T-invariante y que $T_W = I_W$.

(d) Si $V = R(T) \oplus W$ y W es T-invariante, demostrar que $W \subseteq N(T)$. Demostrar también que si V es dimensionalmente finito, entonces $W = N(T)$.

(e) Demostrar que $N(T_W) = N(T) \cap W$ y $R(T_W) = T(W)$.

25. Demostrar la siguiente generalización del Teorema 2.7 para espacios dimensionalmente infinitos: Sean V y W espacios vectoriales y sea β una base para V. Entonces para toda función $f: \beta \to W$ existe únicamente una transformación lineal $T: V \to W$ tal que $T(x) = f(x)$ para toda $x \in \beta$.

26. Una función $T: V \to W$ entre los espacios vectoriales V y W se llama *aditiva* si $T(x + y) = T(x) + T(y)$ para toda $x, y \in V$. Demostrar que si V y W son espacios vectoriales sobre el campo de los números racionales, entonces cualquier función aditiva de V en W es una transformación lineal.

27. Demostrar que existe una función aditiva $T: R \to R$ (como se definió en el Ejercicio 26) que no es lineal. *Sugerencia:* Considérese a R como un espacio vectorial sobre el campo de los números racionales Q. Por el corolario del Teorema 1.15 este espacio vectorial tiene una base β. Sean x y y elementos distintos de β y defínase $f: \beta \to R$ mediante $f(x) = y$, $f(y) = x$ y $f(z) = z$ en cualquier otro caso. Por el Ejercicio 26 existe una transformación lineal $T: R \to R$ donde R se considera como un espacio vectorial sobre Q tal que $T(z) = f(z)$ para toda $z \in \beta$. Entonces T es aditiva pero para $c = y/x$, $T(cx) \neq cT(x)$.

2.2 REPRESENTACION MATRICIAL DE UNA TRANSFORMACION LINEAL

Hasta ahora hemos estudiado las transformaciones lineales examinando sus rangos y sus espacios nulos. Ahora entraremos a uno de los procedimientos de mayor utilidad en el análisis de una transformación lineal sobre un espacio vectorial dimensionalmente finito; la representación de una transformación lineal mediante una matriz. De hecho, desarrollaremos una correspondencia uno-a-uno entre matrices y transformaciones que nos permitirá utilizar las propiedades de una para estudiar las propiedades de la otra.

Necesitaremos primeramente del concepto de una "base ordenada" para un espacio vectorial.

Definición. *Sea* V *un espacio vectorial dimensionalmente finito. Una* base ordenada para V *es una base para* V *establecida con un orden específico; es decir, una base ordenada para* V *es una secuencia finita de elementos de* V *linealmente independientes que generen a* V.

Ejemplo 15. Sea V tal que tenga a $\beta = \{x_1, x_2, x_3\}$ como una base ordenada. Entonces $\gamma = \{x_2, x_1, x_3\}$ es también una base ordenada, pero como bases ordenadas, $\beta \neq \gamma$.

Para el espacio vectorial F^n, llamaremos a $\{e_1, e_2, \ldots, e_n\}$ *la base ordenada estándar para* F^n.

Ahora que contamos con el concepto de base ordenada, seremos capaces de identificar vectores abstractos en un espacio vectorial de n dimensiones con elementos n-dimensionales. Esta identificación será proporcionada mediante el uso de "vectores coordenados" tal como se introducen a continuación.

Definición. *Sea* $\beta = \{x_1, \ldots, x_n\}$ *una base ordenada para un espacio vectorial* V *dimensionalmente finito. Para* $x \in V$ *definimos al* vector coordenado de x relativo a β, *denotado por* $[x]_\beta$, mediante

$$[x]_\beta = \begin{pmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{pmatrix},$$

donde

$$x = \sum_{i=1}^{n} a_i x_i.$$

Nótese que en la definición anterior $[x_i]_\beta = e_i$. Se deja como ejercicio demostrar que la correspondencia $x \to [x]_\beta$ nos proporciona una transformación lineal de V en F^n. Estudiaremos esta transformación con más detalle en la Sección 2.4.

Ejemplo 16. Sean $V = P_2(R)$ y $\beta = \{1, x, x^2\}$. Si $f(x) = 4 + 6x - 7x^2$, entonces

$$[f]_\beta = \begin{pmatrix} 4 \\ 6 \\ -7 \end{pmatrix}.$$

Procedamos ahora con la representación matricial prometida de una transformación lineal. Supóngase que V y W son espacios vectoriales di-

mensionalmente finitos con bases ordenadas $\beta = \{x_1, \ldots, x_n\}$ y $\gamma = \{y_1, \ldots, y_m\}$, respectivamente. Sea $\mathsf{T}\colon \mathsf{V} \to \mathsf{W}$ lineal. Entonces existen escalares únicos $a_{ij} \in F (i = 1, \ldots, m$ y $j = 1, \ldots, n)$ tales que

$$\mathsf{T}(x_j) = \sum_{i=1}^{m} a_{ij} y_j \qquad \text{para } 1 \leq j \leq n.$$

Definición. *Utilizando la notación anterior, llamaremos a la matriz* A *de* m × n, *definida mediante* $A_{ij} = a_{ij}$, *la* matriz que representa a T en las bases ordenadas β y γ *y la escribiremos* $A = [\mathsf{T}]_\beta^\gamma$. *Si* $\mathsf{V} = \mathsf{W}$ *y* $\beta = \gamma$, *escribiremos* $A = [\mathsf{T}]_\beta$.

Nótese que la j-ésima columna de A es sencillamente $[\mathsf{T}(x_j)]_\gamma$. Obsérvese igualmente que del corolario del Teorema 2.7 se concluye que si $\mathsf{U}\colon \mathsf{V} \to \mathsf{W}$ es una transformación lineal tal que $[\mathsf{U}]_\beta^\gamma = [\mathsf{T}]_\beta^\gamma$, entonces $\mathsf{U} = \mathsf{T}$.

Ilustraremos el cálculo de $[\mathsf{T}]_\beta^\gamma$ en los siguientes ejemplos.

Ejemplo 17. Defínase

$$\mathsf{T}\colon \mathsf{P}_3(R) \to \mathsf{P}_2(R) \quad \text{mediante} \quad \mathsf{T}(f) = f'.$$

Sean $\beta = \{1, x, x^2, x^3\}$ y $\gamma = \{1, x, x^2\}$ bases ordenadas para $\mathsf{P}_3(R)$ y $\mathsf{P}_2(R)$, respectivamente. Entonces

$$\begin{aligned} \mathsf{T}(1) &= 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ \mathsf{T}(x) &= 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ \mathsf{T}(x^2) &= 0 \cdot 1 + 2 \cdot x + 0 \cdot x^2 \\ \mathsf{T}(x^3) &= 0 \cdot 1 + 0 \cdot x + 3 \cdot x^2 \end{aligned}$$

Y así se tendrá

$$[\mathsf{T}]_\beta^\gamma = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Nótese que los coeficientes de $\mathsf{T}(x^i)$ cuando se escriben como una combinación de elementos de γ dan los elementos de la columna i-ésima.

Ejemplo 18. Defínase

$$\mathsf{T}\colon \mathsf{R}^2 \to \mathsf{R}^3 \qquad \text{mediante } \mathsf{T}(a_1, a_2) = (a_1 + 3a_2, 0, 2a_1 - 4a_2).$$

Sean β y γ las bases ordenadas estándar para R^2 y R^3, respectivamente Ahora bien,

$$\mathsf{T}(1, 0) = (1, 0, 2) = 1e_1 + 0e_2 + 2e_3$$

y

$$\mathsf{T}(0, 1) = (3, 0, -4) = 3e_1 + 0e_2 - 4e_3.$$

Por lo tanto, matricialmente,

$$[\mathsf{T}]_\beta^\gamma = \begin{pmatrix} 1 & 3 \\ 0 & 0 \\ 2 & -4 \end{pmatrix}.$$

Si hacemos $\gamma' = \{e_3, e_2, e_1\}$, entonces

$$[\mathsf{T}]_\beta^{\gamma'} = \begin{pmatrix} 2 & -4 \\ 0 & 0 \\ 1 & 3 \end{pmatrix}.$$

Ahora que hemos definido un procedimiento para asociar matrices con transformaciones lineales, veremos brevemente que esta asociación "conserva" la adición. Para hacer más explícita esta situación, requeriremos de alguna discusión preliminar sobre la adición de transformaciones lineales.

Definición. *Sean* T, U: V → W *funciones arbitrarias, donde* V *y* W *son espacios vectoriales, y sea* a ∈ F. *Definimos* T + U: V → W *mediante* (T + U)(x) = T(x) + U(x) *para toda* x ∈ V *y* aT: V → W mediante (aT)(x) = aT(x) *para toda* x ∈ V.

Por supuesto, esta es la definición usual de la suma y de la multiplicación por escalares para las funciones. Afortunadamente, sin embargo, tenemos el resultado de que la suma de transformaciones lineales es lineal.

Teorema 2.8. *Sean* V *y* W *espacios vectoriales y sean* T, U: V → W *lineales. Entonces para toda* a ∈ F

(a) aT + U *es lineal.*

(b) *Utilizando las operaciones de suma y de multiplicación por escalares, como se definieron anteriormente, la colección de todas las transformaciones lineales de* V *en* W, *denotada por* $\mathcal{L}$(V, W), *es un espacio vectorial sobre* F.

DEMOSTRACIÓN.

(a) Sean x, $y \in$ V y $c \in F$. Entonces

$$\begin{aligned}(a\mathsf{T} + \mathsf{U})(cx + y) &= a\mathsf{T}(cx + y) + \mathsf{U}(cx + y) \\ &= a[c\mathsf{T}(x) + \mathsf{T}(y)] + c\mathsf{U}(x) + \mathsf{U}(y) \\ &= ac\mathsf{T}(x) + c\mathsf{U}(x) + a\mathsf{T}(y) + \mathsf{U}(y) \\ &= c[a\mathsf{T} + \mathsf{U}](x) + [a\mathsf{T} + \mathsf{U}](y).\end{aligned}$$

Y tenemos que aT + U es lineal.

(b) Observando que T_0, la transformación cero, juega el papel del elemento cero en $\mathcal{L}(V, W)$, es fácil demostrar que $\mathcal{L}(V, W)$ es un espacio vectorial sobre F. ■

En el caso donde $V = W$ escribiremos $\mathcal{L}(V)$ en vez de $\mathcal{L}(V, V)$.

En la sección siguiente veremos una identificación completa de $\mathcal{L}(V, W)$ con el espacio vectorial $M_{m\times n}(F)$, donde n y m son las dimensiones de V y W, respectivamente. Esta identificación se establece fácilmente utilizando el teorema siguiente.

Teorema 2.9. *Sean V y W espacios vectoriales dimensionalmente finitos con bases ordenadas β y γ, respectivamente, y sean $T, U\colon V \to W$ transformaciones lineales. Entonces*

(a) $[T + U]_\beta^\gamma = [T]_\beta^\gamma + [U]_\beta^\gamma$.

(b) $[aT]_\beta^\gamma = a[T]_\beta^\gamma$ *para toda* $a \in F$.

DEMOSTRACIÓN. Sea $\beta = \{x_1, \ldots, x_n\}$ y $\gamma = \{y_1, \ldots, y_m\}$. Existen escalares únicos a_{ij} y b_{ij} en $F(1 \leq i \leq m,\ 1 \leq j \leq n)$ tales que

$$T(x_j) = \sum_{i=1}^{m} a_{ij}y_i \quad \text{y} \quad U(x_j) = \sum_{i=1}^{m} b_{ij}y_i \qquad \text{para } 1 \leq j \leq n.$$

Por lo tanto

$$(T + U)(x_j) = \sum_{i=1}^{m} (a_{ij} + b_{ij})y_i.$$

Y entonces

$$([T + U]_\beta^\gamma)_{ij} = a_{ij} = ([T]_\beta^\gamma + [U]_\beta^\gamma)_{ij}.$$

Así queda demostrado (a) y la demostración de (b) es semejante. ■

Ejemplo 19. Defínase

$$T\colon R^2 \to R^3 \qquad \text{mediante } T(a_1, a_2) = (a_1 + 3a_2, 0, 2a_1, -4a_2)$$

y

$$U\colon R^2 \to R^3 \qquad \text{mediante } U(a_1, a_2) = (a_1 - a_2, 2a_1, 3a_1 + 2a_2).$$

Sean β y γ, respectivamente, bases ordenadas estándar de R^2 y R^3. Entonces

$$[T]_\beta^\gamma = \begin{pmatrix} 1 & 3 \\ 0 & 0 \\ 2 & -4 \end{pmatrix},$$

(tal como se calculó en el Ejemplo 18) y

$$[\mathsf{U}]_\beta^\gamma = \begin{pmatrix} 1 & -1 \\ 2 & 0 \\ 3 & 2 \end{pmatrix}.$$

Si calculamos ahora $\mathsf{T} + \mathsf{U}$ utilizando las definiciones anteriores, obtenemos

$$(\mathsf{T} + \mathsf{U})(a_1, a_2) = (2a_1 + 2a_2, 2a_1, 5a_1 - 2a_2).$$

Entonces

$$[\mathsf{T} + \mathsf{U}]_\beta^\gamma = \begin{pmatrix} 2 & 2 \\ 2 & 0 \\ 5 & -2 \end{pmatrix},$$

que es sencillamente $[\mathsf{T}]_\beta^\gamma + [\mathsf{U}]_\beta^\gamma$, lo que verifica al Teorema 2.9.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas. Para lo siguiente, V y W serán espacios vectoriales dimensionalmente finitos con bases ordenadas β y γ, respectivamente. Suponer que $\mathsf{T}, \mathsf{U}\colon \mathsf{V} \to \mathsf{W}$ son lineales.

   (a) Para cualquier escalar a, $a\mathsf{T} + \mathsf{U}$ es una transformación lineal de V en W.
   (b) $[\mathsf{T}]_\beta^\gamma = [\mathsf{U}]_\beta^\gamma$ implica que $\mathsf{T} = \mathsf{U}$.
   (c) Si $m = \dim(\mathsf{V})$ y $n = \dim(\mathsf{W})$, entonces $[\mathsf{T}]_\beta^\gamma$ es una matriz de $m \times n$.
   (d) $[\mathsf{T} + \mathsf{U}]_\beta^\gamma = [\mathsf{T}]_\beta^\gamma + [\mathsf{U}]_\beta^\gamma$.
   (e) $\mathcal{L}(\mathsf{V}, \mathsf{W})$ es un espacio vectorial.
   (f) $\mathcal{L}(\mathsf{V}, \mathsf{W}) = \mathcal{L}(\mathsf{W}, \mathsf{V})$.

2. Sean β y γ las bases ordenadas estándar para R^n y R^m, respectivamente. Para las siguientes transformaciones $\mathsf{T}\colon \mathsf{R}^n \to \mathsf{R}^m$, calcular $[\mathsf{T}]_\beta^\gamma$.

   (a) $\mathsf{T}\colon \mathsf{R}^2 \to \mathsf{R}^3$ definida mediante $\mathsf{T}(a_1, a_2) = (2a_1 - a_2, 3a_1 + 4a_2, a_1)$.
   (b) $\mathsf{T}\colon \mathsf{R}^3 \to \mathsf{R}^2$ definida mediante $\mathsf{T}(a_1, a_2, a_3) = (2a_1 + 3a_2 - a_3, a_1 + a_3)$.
   (c) $\mathsf{T}\colon \mathsf{R}^3 \to R$ definida mediante $\mathsf{T}(a_1, a_2, a_3) = 2a_1 + a_2 - 3a_3$.

3. Sea $\mathsf{T}\colon \mathsf{R}^2 \to \mathsf{R}^3$ definida como $\mathsf{T}(a_1, a_2) = (a_1 - a_2, a_1, 2a_1 + a_2)$. Sea β la base ordenada estándar para R^2 y $\gamma = \{(1, 1, 0), (0, 1, 1), (2, 2, 3)\}$. Calcular $[\mathsf{T}]_\beta^\gamma$. Si $\alpha = \{(1, 2), (2, 3)\}$, calcular $[\mathsf{T}]_\alpha^\gamma$.

4. Defínase

$$\mathsf{T}\colon\ \mathsf{M}_{2\times 2}(R) \to \mathsf{P}_2(R) \quad \text{mediante} \quad \mathsf{T}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = (a+b) + (2d)x + bx^2.$$

Sea

$$\beta = \left\{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right\} \quad \text{y} \quad \gamma = \{1, x, x^2\}.$$

Calcular $[\mathsf{T}]_\beta^\gamma$.

5. Para los siguientes incisos, sean

$$\alpha = \left\{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right\},$$

$$\beta = \{1, x, x\}^2,$$

y

$$\gamma = \{1\}.$$

(a) Definida $\mathsf{T}\colon\ \mathsf{M}_{2\times 2}(F) \to \mathsf{M}_{2\times 2}(F)$ mediante $\mathsf{T}(A) = A^t$, calcular $[\mathsf{T}]_\alpha$.

(b) Definida

$$\mathsf{T}\colon\ \mathsf{P}_2(R) \to \mathsf{M}_{2\times 2}(R) \quad \text{mediante} \quad \mathsf{T}(f) = \begin{pmatrix} f'(0) & 2f(1) \\ 0 & f''(3) \end{pmatrix}$$

donde $'$ significa diferenciación, calcular $[\mathsf{T}]_\beta^\alpha$.

(c) Definida $\mathsf{T}\colon\ \mathsf{M}_{2\times 2}(F) \to F$ mediante $\mathsf{T}(A) = \operatorname{tr}(A)$, calcular $[\mathsf{T}]_\alpha^\gamma$.

(d) Definida $\mathsf{T}\colon\ \mathsf{P}_2(R) \to R$ mediante $\mathsf{T}(f) = f(2)$, calcular $[\mathsf{T}]_\beta^\gamma$.

(e) Si

$$A = \begin{pmatrix} 1 & -2 \\ 0 & 4 \end{pmatrix},$$

calcular $[A]_\alpha$.

(f) Si $f(x) = 3 - 6x + x^2$, calcular $[f]_\beta$.

(g) Si $a \in F$, calcular $[a]_\gamma$.

6. Demostrar el inciso (b) del Teorema 2.9.

7.* Sea V un espacio vectorial n-dimensional con una base ordenada β. Definiendo a $\mathsf{T}\colon\ \mathsf{V} \to \mathsf{F}^n$ mediante $\mathsf{T}(x) = [x]_\beta$, demostrar que T es lineal.

8. Sea V el espacio vectorial de los números complejos sobre el campo R. Si $\mathsf{T}\colon\ \mathsf{V} \to \mathsf{V}$ queda definida mediante $\mathsf{T}(z) = \bar{z}$ donde $\bar{z}$ es el complejo conjugado de z, demostrar que T es lineal y calcular $[\mathsf{T}]_\beta$, donde $\beta = \{1, i\}$. Mostrar que T no es lineal si V se considera como un espacio vectorial sobre el campo C.

9. Sea V un espacio vectorial con la base ordenada $\beta = \{x_1, \ldots, x_n\}$. Defínase a $x_0 = 0$. De acuerdo con el Teorema 2.7 debe existir una transformación lineal T: V → V definida mediante $\mathsf{T}(x_j) = x_j + x_{j-1}$ para $j = 1, \ldots, n$. Calcular $[\mathsf{T}]_\beta$.

10. Sea V un espacio vectorial n-dimensional y sea T: V → V una transformación lineal. Supóngase que W es un subespacio T-invariante de V (ver el Ejercicio 24 de la Sección 2.1) de dimensión k. Demostrar que existe una base β para V tal que $[\mathsf{T}]_\beta$ tiene la forma

$$\begin{bmatrix} A & B \\ O & C \end{bmatrix},$$

donde A es una matriz de $k \times k$ y O es una matriz cero de $(n - k) \times k$.

11. Sea V un espacio vectorial de dimensión finita y sea T una proyección sobre un subespacio W de V. Escoger una base ordenada adecuada β para V tal que $[\mathsf{T}]_\beta$ sea una matriz diagonal.

12. Sean V y W espacios vectoriales y sean T y U transformaciones lineales no nulas de V en W. Si $\mathsf{R}(\mathsf{T}) \cap \mathsf{R}(\mathsf{U}) = \{0\}$, demostrar que $\{\mathsf{T}, \mathsf{U}\}$ es un subconjunto de $\mathcal{L}(\mathsf{V}, \mathsf{W})$ linealmente independiente.

13. Sea $\mathsf{V} = \mathsf{P}(R)$, y para $j \geq 0$ defínase a T_j: V → V mediante $\mathsf{T}_j(f) = f^{(j)}$, donde $f^{(j)}$ sea la j-ésima derivada de f. Para cualquier entero positivo n, demostrar que $\{\mathsf{T}_1, \mathsf{T}_2, \ldots, \mathsf{T}_n\}$ es un subconjunto de $\mathcal{L}(\mathsf{V})$ linealmente independiente.

14. Sean V y W espacios vectoriales y sea S subconjunto de V. Defínase $S^0 = \{\mathsf{T} \in \mathcal{L}(\mathsf{V}, \mathsf{W})\colon \mathsf{T}(x) = 0 \text{ para toda } x \in S\}$. Demostrar

(a) S^0 es un subespacio de $\mathcal{L}(\mathsf{V}, \mathsf{W})$.
(b) Si S_1 y S_2 son subconjuntos de V y $S_1 \subseteq S_2$, entonces $S_2^0 \subseteq S_1^0$.
(c) Si V_1 y V_2 son subespacios de V, entonces $(\mathsf{V}_1 + \mathsf{V}_2)^0 = \mathsf{V}_1^0 \cap \mathsf{V}_2^0$.

15. Sean V y W espacios vectoriales dimensionalmente finitos y sea T: V → W lineal. Supóngase que $\dim(\mathsf{V}) = \dim(\mathsf{W})$. Encontrar bases ordenadas β y γ para V y W, respectivamente, tales que $[\mathsf{T}]_\beta^\gamma$ sea una matriz diagonal.

2.3 COMPOSICION DE TRANSFORMACIONES LINEALES Y MULTIPLICACION DE MATRICES

En la Sección 2.2 aprendimos cómo asociar una matriz con un transformación lineal de tal modo que las sumas de matrices quedaban asociadas con las correspondientes sumas de transformaciones. Ahora surge la pregunta sobre cómo se relaciona la representación matricial de una composición de transformaciones lineales con las representaciones matriciales de

cada una de las transformaciones lineales asociadas. El intento de responder a esta interrogante nos conducirá a una definición de multiplicación de matrices. Utilizaremos la notación UT para la composición de las transformaciones lineales U y T, contrastando con $g \circ f$ para funciones arbitrarias g y f. Específicamente, tenemos la definición siguiente.

Definición. *Sean* V, W *y* Z *espacios vectoriales y sean* $\mathsf{T}\colon \mathsf{V} \to \mathsf{W}$ *y* $\mathsf{U}\colon \mathsf{W} \to \mathsf{Z}$ *lineales. Definamos* $\mathsf{UT}\colon \mathsf{V} \to \mathsf{Z}$ *mediante* $(\mathsf{UT})(\mathsf{x}) = \mathsf{U}(\mathsf{T}(\mathsf{x}))$ *para toda* $\mathsf{x} \in \mathsf{V}$.

Nuestro primer resultado muestra que la composición de transformaciones lineales es lineal.

Teorema 2.10. *Sean* V, W *y* Z *espacios vectoriales y* $\mathsf{T}\colon \mathsf{V} \to \mathsf{W}$ *y* $\mathsf{U}\colon \mathsf{W} \to \mathsf{Z}$ *lineales. Entonces* $\mathsf{UT}\colon \mathsf{V} \to \mathsf{Z}$ *es lineal.*

DEMOSTRACIÓN. Sean $x, y \in \mathsf{V}$ y $a \in F$. Entonces

$$\begin{aligned} \mathsf{UT}(ax + y) &= \mathsf{U}(\mathsf{T}(ax + y)) = \mathsf{U}(a\mathsf{T}(x) + \mathsf{T}(y)) \\ &= a\mathsf{U}(\mathsf{T}(x)) + \mathsf{U}(\mathsf{T}(y)) = a(\mathsf{UT})(x) + \mathsf{UT}(y). \quad \blacksquare \end{aligned}$$

El siguiente teorema enuncia algunas de las propiedades de la composición de transformaciones lineales.

Teorema 2.11. *Sea* V *un espacio vectorial. Sean* $\mathsf{T}, \mathsf{U}_1, \mathsf{U}_2 \in \mathcal{L}(\mathsf{V})$. *Entonces*

(a) $\mathsf{T}(\mathsf{U}_1 + \mathsf{U}_2) = \mathsf{TU}_1 + \mathsf{TU}_2$ y $(\mathsf{U}_1 + \mathsf{U}_2)\mathsf{T} = \mathsf{U}_1\mathsf{T} + \mathsf{U}_2\mathsf{T}$.
(b) $\mathsf{T}(\mathsf{U}_1\mathsf{U}_2) = (\mathsf{TU}_1)\mathsf{U}_2$.
(c) $\mathsf{TI} = \mathsf{IT} = \mathsf{T}$.
(d) $\mathsf{a}(\mathsf{U}_1\mathsf{U}_2) = (\mathsf{aU}_1)(\mathsf{U}_2 = \mathsf{U}_1(\mathsf{aU}_2)$ *para toda* $\mathsf{a} \in \mathsf{F}$.

DEMOSTRACIÓN. Ejercicio.

Estamos ahora en posición de definir el producto AB de dos matrices A y B. Por el Teorema 2.9, parece razonable requerir por analogía que si $A = [\mathsf{U}]_\beta^\gamma$ y $B = [\mathsf{T}]_\alpha^\beta$, donde $\mathsf{T}\colon \mathsf{V} \to \mathsf{W}$ y $\mathsf{U}\colon \mathsf{W} \to \mathsf{Z}$, entonces $AB = [\mathsf{UT}]_\alpha^\gamma$.

Ahora sean T, U, A y B como anteriormente y sean $\alpha = \{x_1, \ldots, x_n\}$, $\beta = \{y_1, \ldots, y_m\}$ y $\gamma = \{z_1, \ldots, z_p\}$ bases ordenadas para V, W y Z, respectivamente. Para $1 \le j \le n$ tendremos

$$\begin{aligned} (\mathsf{UT})(x_j) = \mathsf{U}(\mathsf{T}(x_j)) &= \mathsf{U}\left(\sum_{k=1}^{m} B_{kj}y_k\right) = \sum_{k=1}^{m} B_{kj}\mathsf{U}(y_k) \\ &= \sum_{k=1}^{m} B_{kj}\left(\sum_{i=1}^{p} A_{ik}z_i\right) = \sum_{i=1}^{p}\left(\sum_{k=1}^{m} A_{ik}B_{kj}\right)z_i \\ &= \sum_{i=1}^{p} C_{ij}z_i, \end{aligned}$$

donde

$$C_{ij} = \sum_{k=1}^{m} A_{ik}B_{kj}.$$

Este cálculo sugiere la siguiente definición de multiplicación de matrices.

Definición. *Sean* A *una matriz de* m × n *y* B *una matriz de* n × p. *Definimos el producto de* A *por* B, *denotado por* AB, *como la matriz de* m × p *tal que*

$$(\mathrm{AB})_{ij} = \sum_{k=1}^{n} \mathrm{A}_{ik}\mathrm{B}_{kj} \qquad \textit{para } 1 \leq \mathrm{i} \leq \mathrm{m}, \quad 1 \leq \mathrm{j} \leq \mathrm{p}.$$

Nótese que $(AB)_{ij}$ es la suma de los productos de los elementos correspondientes al i-ésimo renglón de A y a la j-ésima columna de B.

Al final de esta sección el lector verá algunas aplicaciones interesantes de esta definición.

El lector debe observar que para que el producto AB quede definido, existen restricciones en cuanto a las dimensiones relativas de A y B. El siguiente dispositivo simbólico puede ser útil: "$(m \times n) \cdot (n \times p) = (m \times p)$"; esto es, para que el producto AB esté definido, las dos dimensiones "interiores" deben ser iguales y las dos dimensiones "exteriores" dan el tamaño del producto.

Ejemplo 20.

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 4 & -1 \end{pmatrix}\begin{pmatrix} 4 \\ 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 4 + 4 + 5 \\ 0 + 8 - 5 \end{pmatrix} = \begin{pmatrix} 13 \\ 3 \end{pmatrix}$$

Nótese de nuevo la relación simbólica $(2 \times 3) \cdot (3 \times 1) = 2 \times 1$.

Como en el caso de la composición de funciones, tenemos que el producto de matrices no es conmutativo. Considérese los dos productos siguientes.

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 3 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 7 \\ 1 & 2 \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} 0 & 3 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 3 \\ 1 & 4 \end{pmatrix}.$$

Donde vemos que aun cuando ambos productos matriciales AB y BA están definidos, no es necesariamente cierto que $AB = BA$.

Recordando la definición de transpuesta de una matriz dada en la Sección 1.3, demostraremos que si A es una matriz de $m \times n$ y B es una matriz de $n \times p$, entonces $(AB)^t = B^tA^t$. Como

$$(AB)^t_{ij} = (AB)_{ji} = \sum_{k=1}^{n} A_{jk}B_{ki}$$

y

$$(B^tA^t)_{ij} = \sum_{k=1}^{n} (B^t)_{ik}(A^t)_{kj} = \sum_{k=1}^{n} B_{ki}A_{jk},$$

ya queda demostrado. Por lo tanto, la transpuesta de un producto es igual al producto de las transpuestas, *cambiando el orden* de las matrices que se multiplican.

El teorema siguiente es una consecuencia inmediata de nuestra definición del producto de matrices.

Teorema 2.12. *Sean* V, W *y* Z *espacios vectoriales dimensionalmente finitos con bases ordenadas* α, β *y* γ, *respectivamente. Sean* T: V ⟶ W *y* U: W ⟶ Z *transformaciones lineales. Entonces*

$$[\mathsf{UT}]_\alpha^\gamma = [\mathsf{U}]_\beta^\gamma[\mathsf{T}]_\alpha^\beta.$$

Corolario. *Sea* V *un espacio vectorial dimensionalmente finito con una base ordenada* β. *Sean* T, U ∈ $\mathcal{L}$(V). *Entonces* $[\mathsf{UT}]_\beta = [\mathsf{U}]_\beta[\mathsf{T}]_\beta$.

Ilustraremos lo anterior con el siguiente ejemplo.

Ejemplo 21. Defínase

$$\mathsf{U}: \ \mathsf{P}_3(R) \to \mathsf{P}_2(R) \quad \text{mediante} \quad \mathsf{U}(f) = f'$$

como en el Ejemplo 17. Defínase

$$\mathsf{T}: \ \mathsf{P}_2(R) \to \mathsf{P}_3(R) \quad \text{mediante} \quad \mathsf{T}(f)(x) = \int_0^x f(t)dt.$$

Sean $\alpha = \{1, x, x^2, x^3\}$ y $\beta = \{1, x, x^2\}$. Se tiene claramente que UT = I. Para ilustrar el Teorema 2.12, obsérvese que

$$[\mathsf{UT}]_\beta = [\mathsf{U}]_\alpha^\beta[\mathsf{T}]_\beta^\alpha = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = [\mathsf{I}]_\beta.$$

La matriz diagonal de 3×3 anterior se llama "matriz identidad" y se define a continuación junto con una notación de gran utilidad, la "delta de Kronecker".

Definiciones. *Definimos la* delta de Kronecker δ_{ij} *mediante* $\delta_{ij} = 1$ *si* $i = j$ *y* $\delta_{ij} = 0$ *si* $i \neq j$, *y la* matriz identidad de $n \times n$, I_n, *mediante* $(I_n)_{ij} = \delta_{ij}$.

Así

$$I_1 = (1), \qquad I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{y} \quad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Veremos en el teorema siguiente que la matriz identidad actúa como un elemento unitario en $M_{n\times n}(F)$. Cuando el contexto sea lo suficientemente claro, omitiremos algunas veces el subíndice n de I_n.

Teorema 2.13. *Para cualquier matriz* A *de* n $\times$ n *tenemos que* $I_nA = AI_n = A$, *además, si* V *es un espacio vectorial dimensionalmente finito de dimensión* n *con una base ordenada* β, *entonces* $[I_V]_\beta = I_n$.

DEMOSTRACIÓN.

$$(I_nA)_{ij} = \sum_{k=1}^{n} (I_n)_{ik}A_{kj} = \sum_{k=1}^{n} \delta_{ik}A_{kj} = A_{ij}.$$

Por lo tanto, $I_nA = A$. De la misma manera $AI_n = A$. Sea $\beta = \{x_1, \ldots, x_n\}$. Entonces, para cada j tenemos

$$I_V(x_j) = x_j = \sum_{i=1}^{n} \delta_{ij}x_i.$$

Por lo tanto $[I_V]_\beta = I_n$. ■

Para cualquier matriz A de $n \times n$ definiremos $A^2 = AA$, $A^3 = A^2A$, y en general $A^k = A^{k-1}A$ para $k = 2, 3, \ldots$ Definiremos $A^0 = I_n$.

Con esta notación vemos que si

$$A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},$$

entonces $A^2 = O$ (la matriz nula) aun cuando $A \neq O$, y vemos que la propiedad de eliminación (cancelación) para campos no es válida para las matrices. El siguiente teorema muestra, sin embargo, que la multiplicación de matrices es distributiva respecto a la suma.

Teorema 2.14. *Sea* A *una matriz de* m $\times$ n *y sean* B *y* C *matrices de* n $\times$ p. *Entonces*

$$A(B + C) = AB + AC,$$

y para cualquier escalar a

$$a(AB) = (aA)B = A(aB).$$

DEMOSTRACIÓN.

$$[A(B + C)]_{ij} = \sum_{k=1}^{n} A_{ik}(B + C)_{kj} = \sum_{k=1}^{n} A_{ik}(B_{kj} + C_{kj})$$

$$= \sum_{k=1}^{n} (A_{ik}B_{kj} + A_{ik}C_{kj}) = \sum_{k=1}^{n} A_{ik}B_{kj} + \sum_{k=1}^{n} A_{ik}C_{kj}$$

$$= (AB)_{ij} + (AC)_{ij} = [AB + AC]_{ik}.$$

El resto de la demostración se deja como ejercicio. ■

Corolario. *Sea* A *una matriz de* m × n, *y sean* $B_1, \ldots, B_k$ *matrices de* n × p *y* $a_1, \ldots, a_k \in F$. *Entonces*

$$A\left(\sum_{i=1}^{k} a_i B_i\right) = \sum_{i=1}^{k} a_i A B_i.$$

DEMOSTRACIÓN. Ejercicio.

Si A es una matriz de $m \times n$, escribiremos a veces $A = (A^1, \ldots, A^n)$, donde A^j es la columna j-ésima

$$\begin{pmatrix} A_{1j} \\ A_{2j} \\ \cdot \\ \cdot \\ \cdot \\ A_{mj} \end{pmatrix}$$

de la matriz A.

Para el teorema siguiente, e_j representa la j-ésima columna de I_p.

Teorema 2.15. *Sea* A *una matriz de* m × n *y* B *una matriz de* n × p. *Entonces*

(a) $(AB)^j = AB^j$.
(b) $B^j = Be_j$.

DEMOSTRACIÓN.

$$(AB)^j = \begin{pmatrix} (AB)_{1j} \\ \cdot \\ \cdot \\ \cdot \\ (AB)_{mj} \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{n} A_{1k}B_{kj} \\ \cdot \\ \cdot \\ \cdot \\ \sum_{k=1}^{n} A_{mk}B_{kj} \end{pmatrix} = A\begin{pmatrix} B_{1j} \\ \cdot \\ \cdot \\ B_{nj} \end{pmatrix} = AB^j.$$

Por lo tanto, (a) queda demostrado. La demostración de (b) se deja como ejercicio. ∎

El resultado siguiente justificará mucho de nuestro trabajo anterior; utilizará la representación matricial de una transformación lineal y el producto de matrices para evaluar la transformación en cualquier vector dado.

Teorema 2.16. *Sean* V *y* W *espacios vectoriales dimensionalmente finitos que tienen bases ordenadas* β *y* γ, *respectivamente, y sea* T: V → W *una transformación lineal. Entonces, para toda* $x \in V$ *tenemos*

$$[T(x)]_\gamma = [T]_\beta^\gamma [x]_\beta.$$

DEMOSTRACIÓN. Sea $\beta = \{x_1, \ldots, x_n\}$ y sea $A = [\mathrm{T}]_\beta^\gamma$. Como

$$\left[\mathrm{T}\left(\sum_{i=1}^{n} a_i x_i\right)\right]_\gamma = \left[\sum_{i=1}^{n} a_i \mathrm{T}(x_i)\right]_\gamma = \sum_{i=1}^{n} a_i [\mathrm{T}(x_i)]_\gamma$$

por el Ejercicio 7 de la Sección 2.2, según el corolario del Teorema 2.14 es suficiente demostrar el teorema para $x = x_j (1 \leq j \leq n)$. Pero esto se sigue de la definición de A y del Teorema 2.15 puesto que

$$[\mathrm{T}(x_j)]_\gamma = A^j = Ae_j = A[x_j]_\beta = [\mathrm{T}]_\beta^\gamma [x_j]_\beta. \quad \blacksquare$$

Ejemplo 22. Sea T: $\mathrm{P}_3(R) \to \mathrm{P}_2(R)$ definida mediante $\mathrm{T}(f) = f'$, y sean $\beta = \{1, x, x^2, x^3\}$ y $\gamma = \{1, x, x^2\}$ bases ordenadas para $\mathrm{P}_3(R)$ y $\mathrm{P}_2(R)$, respectivamente. Si $A = [\mathrm{T}]_\beta^\gamma$, entonces tenemos del Ejemplo 17 que

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Ilustraremos el Teorema 2.16 verificando que $[\mathrm{T}(p)]_\gamma = [\mathrm{T}]_\beta^\gamma [p]_\beta$ donde $p \in \mathrm{P}_3(R)$ es el polinomio $p(x) = 2 - 4x + x^2 + 3x^3$. Sea $q = \mathrm{T}(p)$; entonces $q(x) = p'(x) = -4 + 2x + 9x^2$. Por lo tanto

$$[\mathrm{T}(p)]_\gamma = [q]_\gamma = \begin{pmatrix} -4 \\ 2 \\ 9 \end{pmatrix}.$$

Pero también

$$[\mathrm{T}]_\beta^\gamma [p]_\beta = A[p]_\beta = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ -4 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} -4 \\ 2 \\ 9 \end{pmatrix}.$$

Completaremos esta sección con la introducción de la "transformación de multiplicación por la izquierda" L_A, donde A es una matriz de $m \times n$. Esta transformación es probablemente la herramienta más importante para transferir propiedades sobre transformaciones a propiedades semejantes sobre matrices y viceversa. Por ejemplo, la utilizaremos para demostrar que el producto de matrices es asociativo.

Definición. *Sea* A *una matriz de* m × n *con elementos de un campo* F. *Denotamos por* L_A *al mapeo* L_A: $\mathrm{F}^n \to \mathrm{F}^m$ *definido por* $\mathrm{L}_A(x) = Ax$ *(el producto matricial de* A *por* x*) para cada vector columna* $x \in \mathrm{F}^n$. *Llamamos a* L_A *una* transformación de multiplicación por la izquierda.

Ejemplo 23. Sea

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

Si

$$x = \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix},$$

entonces

$$\mathsf{L}_A(x) = Ax = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix} = \begin{pmatrix} 6 \\ 1 \end{pmatrix}.$$

Veremos en el teorema que sigue que L_A no únicamente es lineal, sino, de hecho, tiene otras muchas propiedades de gran utilidad. Estas propiedades son todas muy naturales y por lo tanto son fáciles de recordar.

Teorema 2.17. *Sea* A *una matriz de* $m \times n$ *con elementos de* F. *Entonces la transformación* $\mathsf{L}_A\colon \mathsf{F}^n \to \mathsf{F}^m$ *es lineal. Además, si* B *es cualquiera otra matriz de* $m \times n$ *(con elementos de* F*), tenemos las siguientes propiedades.*

(a) $[\mathsf{L}_A]_\beta^\gamma = A$, *donde* B *y* γ *son las bases ordenadas estándar para* F^n *y* F^m, *respectivamente.*
(b) $\mathsf{L}_A = \mathsf{L}_B$ *si y sólo si* $A = B$.
(c) $\mathsf{L}_{A+B} = \mathsf{L}_A + \mathsf{L}_B$ *y* $\mathsf{L}_{aA} = a\mathsf{L}_A$ *para toda* $a \in F$.
(d) *Si* $\mathsf{T}\colon \mathsf{F}^n \to \mathsf{F}^m$ *es lineal, entonces existe una matriz única* C *de* $m \times n$ *tal que* $\mathsf{T} = \mathsf{L}_C$.
(e) *Si* E *es una matriz de* $n \times p$, *entonces* $\mathsf{L}_{AE} = \mathsf{L}_A\mathsf{L}_E$.
(f) *Si* $m = n$, *entonces* $I_{I_n} = \mathsf{I}_{\mathsf{F}^n}$.

DEMOSTRACIÓN. El hecho de que L_A sea lineal se deriva directamente del Teorema 2.14 y su corolario.

(a) La columna j-ésima de $[\mathsf{L}_A]_\beta^\gamma$ es igual a $\mathsf{L}_A(e_j)$. Pero $\mathsf{L}_A(e_j) = Ae_j = A^j$, y entonces $[\mathsf{L}_A]_\beta^\gamma = A$.
(b) Si $\mathsf{L}_A = \mathsf{L}_B$ entonces podemos utilizar (a) para escribir $A = [\mathsf{L}_A]_\beta^\gamma = [\mathsf{L}_B]_\beta^\gamma = B$. Por lo tanto $A = B$. La prueba de la proposición recíproca es trivial.

La demostración de (c) le corresponde al lector.

(d) Sea $C = [\mathsf{T}]_\beta^\gamma$. En virtud del Teorema 2.16 tenemos que $[\mathsf{T}(x)]_\gamma = [\mathsf{T}]_\beta^\gamma[x]_\beta$, o bien $\mathsf{T}(x) = Cx = \mathsf{L}_C(x)$ para toda x. Entonces $\mathsf{T} = \mathsf{L}_C$. La unicidad de C se deduce de (b).

(e) Para cualquier j tenemos $\mathsf{L}_{AE}(e_j) = (AE)e_j = (AE)^j = AE^j = A(Ee_j) = \mathsf{L}_A(Ee_j) = \mathsf{L}_A(\mathsf{L}_E(e_j)) = (\mathsf{L}_A\mathsf{L}_E)(e_j)$. De donde $\mathsf{L}_{AE} = \mathsf{L}_A\mathsf{L}_E$ por el Teorema 2.7.

La demostración de (f) le corresponde al lector. ■

Utilizaremos este tipo de transformaciones para establecer una importante propiedad sobre matrices.

Teorema 2.18. *Sean* A, B *y* C *matrices tales que* A(BC) *está definido. Entonces* (AB)C *queda definido y* A(BC) = (AB)C; *esto significa que el producto de matrices es asociativo.*

DEMOSTRACIÓN. Se deja al lector demostrar que $(AB)C$ está definido. Utilizando el inciso (e) del Teorema 2.17 y la asociatividad de una composición de funciones, tenemos

$$\mathsf{L}_{A(BC)} = \mathsf{L}_A\mathsf{L}_{BC} = \mathsf{L}_A(\mathsf{L}_B\mathsf{L}_C) = (\mathsf{L}_A\mathsf{L}_B)\mathsf{L}_C = \mathsf{L}_{AB}\mathsf{L}_C = \mathsf{L}_{(AB)C}.$$

Así, del inciso (b) del Teorema 2.17 tenemos $A(BC) = (AB)C$. ■

Es innecesario decir que este teorema podría demostrarse directamente a partir de la definición del producto matricial. La demostración anterior, sin embargo, proporciona un prototipo de muchos otros argumentos que utilizan las relaciones entre transformaciones lineales y matrices.

Una aplicación

Una grande y variada colección de aplicaciones interesantes surge en relación con unas matrices especiales llamadas "matrices incidentes". Una *matriz incidente* es una matriz cuadrada en donde todos los elementos son ceros o unos y, por conveniencia, todos los elementos de la diagonal principal son cero. Si tenemos una cierta relación entre un grupo de n objetos que denotaremos por $1, 2, \ldots, n$, entonces definimos la matriz de incidencia asociada A mediante: $A_{ij} = 1$ si i está relacionado con j y $A_{ij} = 0$ en cualquier otro caso.

Para hacer las cosas concretas, supongamos que tenemos cuatro personas, cada una de las cuales posee un dispositivo de comunicación. Si la relación entre este grupo es "puede transmitir a", entonces $A_{ij} = 1$ si i puede mandar un mensaje a j y $A_{ij} = 0$ en cualquier otro caso. Supóngase que

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$$

Entonces como $A_{34} = 1$ y $A_{14} = 0$ vemos que la persona 3 puede transmitir a 4 pero 1 no puede transmitir a 4.

Obtendremos una interpretación interesante de los elementos de A^2. Considérese por ejemplo,

$$(A^2)_{31} = A_{31}A_{11} + A_{32}A_{21} + A_{33}A_{31} + A_{34}A_{41}.$$

Nótese que cualquiera de los términos $A_{3k}A_{k1}$ será igual a 1 si y sólo si A_{3k} y A_{k1} son iguales a 1 esto es, si y sólo si 3 puede transmitir a k y k puede transmitir a 1. Así $(A^2)_{31}$ da el número de maneras en las que 3 puede transmitir a 1 en dos *etapas* (o en un *relevo).* Como

$$A^2 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 2 & 0 & 0 \\ 2 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix},$$

vemos que 3 puede transmitir a 1 de dos maneras con un relevo. En general $(A + A^2 + \ldots + A^n)_{ij}$ es el número de maneras en las que i puede transmitir a j en un máximo de n etapas.

Una colección máxima de tres o más personas con la propiedad de que cualquier par de ellas transmita recíprocamente de una a otra se llama una *cliqué.* El problema de la determinación de las cliqués parece a primera vista demasiado difícil. Sin embargo, si se define una nueva matriz B por $B_{ij} = 1$ si i y j pueden transmitirse de una a otra y de lo contrario $B_{ij} = 0$, entonces puede mostrarse (ver Ejercicio 16) que la persona i pertenece a una cliqué si y sólo si $(B^3)_{ii} > 0$. Por ejemplo, supóngase que la matriz de incidencia asociada con alguna relación es

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

Para determinar qué personas pertenecen a las cliqués, formamos una matriz B como la anterior y calculamos B^3. En este caso

$$B = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \quad \text{y} \quad B^3 = \begin{pmatrix} 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 \\ 0 & 4 & 0 & 4 \\ 4 & 0 & 4 & 0 \end{pmatrix}.$$

Como todos los elementos de la diagonal de B^3 son ceros, concluimos que no hay cliqués en esta relación.

Nuestro ejemplo final acerca del uso de matrices de incidencia se relaciona con el concepto de dominancia. Una relación entre un grupo de

personas se llama *relación de dominancia* si la matriz de incidencia asociada A tiene la propiedad de que $A_{ij} = 1$ si y sólo si $A_{ji} = 0$ para toda i y toda j, esto es, dadas dos personas cualesquiera, exactamente una de ellas domina (o, utilizando la terminología de nuestro primer ejemplo, puede transmitir un mensaje) a la otra. Para tal relación, puede demostrarse (ver Ejercicio 18) que la matriz $A + A^2$ tiene un renglón (columna) que contiene elementos positivos en todas las posiciones excepto en la diagonal principal. En otras palabras, existe al menos una persona que domina a (es dominada por) todas las demás en una o dos etapas. De hecho, puede demostrarse que cualquier persona que domina a (es dominada por) el mayor número de personas en la primera etapa cumple con esta propiedad. Considérese, por ejemplo, a la matriz

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix}.$$

El lector deberá verificar que esta matriz corresponde a una relación de dominancia. Ahora bien,

$$A + A^2 = \begin{pmatrix} 0 & 2 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 2 & 0 & 2 & 1 \\ 1 & 2 & 2 & 0 & 1 \\ 2 & 2 & 2 & 2 & 0 \end{pmatrix}.$$

Entonces, las personas 1, 3, 4 y 5 dominan a (pueden mandar mensajes a) todas las demás en a lo más dos etapas, mientras que las personas 1, 2, 3 y 4 son dominadas por (pueden recibir mensajes de) todas las demás en a lo más dos etapas.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas. Para lo que sigue, V, W y Z son espacios vectoriales con bases ordenadas (finitas) α, β y γ, respectivamente; T: $\mathsf{V} \to \mathsf{W}$ y U: $\mathsf{W} \to \mathsf{Z}$ son ambas lineales; y A y B son matrices
   (a) $[\mathsf{UT}]_\alpha^\gamma = [\mathsf{U}]_\beta^\gamma [\mathsf{T}]_\gamma^\beta$.
   (b) $[\mathsf{T}(x)]_\beta = [\mathsf{T}]_\alpha^\beta [x]_\alpha$ para toda $x \in \mathsf{V}$.
   (c) $[\mathsf{U}(y)]_\beta = [\mathsf{U}]_\alpha^\beta [y]_\beta$ para toda $y \in \mathsf{W}$.
   (d) $[\mathsf{I_V}]_\alpha = I$.
   (e) $[\mathsf{T}^2]_\alpha^\beta = ([\mathsf{T}]_\alpha^\beta)^2$.

(f) $A^2 = I$ implica que $A = I$ o bien $A = -I$.
(g) $\mathsf{T} = \mathsf{L}_A$ para alguna matriz A.
(h) $A^2 = O$ implica que $A = O$ donde O denota la matriz nula.
(i) $\mathsf{L}_{A+B} = \mathsf{L}_A + \mathsf{L}_B$.
(j) Si A es cuadrada y $A_{ij} = \delta_{ij}$ para toda i y j, entonces $A = I$.

2. Sean

$$A = \begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 & -3 \\ 4 & 1 & 2 \end{pmatrix},$$

$$C = \begin{pmatrix} 1 & 1 & 4 \\ -1 & -2 & 0 \end{pmatrix}, \quad \text{y} \quad D = \begin{pmatrix} 2 \\ -2 \\ 3 \end{pmatrix}.$$

Calcular $A(2B + 3C)$, $(AB)D$ y $A(BD)$.

3. Sea $g(x) = 3 + x$. Defínase

$$\mathsf{T}\colon\ \mathsf{P}_2(R) \to \mathsf{P}_2(R) \quad \text{para} \quad \mathsf{T}(f) = f'g + 2f.$$

$$\mathsf{U}\colon\ \mathsf{P}_2(R) \to \mathsf{R}^3 \quad \text{para} \quad \mathsf{U}(a + bx + cx^2) = (a + b,\ c,\ a - b).$$

Sean $\beta = \{1, x, x^2\}$ y $\gamma = \{e_1, e_2, e_3\}$.

(a) Calcular directamente $[\mathsf{U}]_\beta^\gamma$, $[\mathsf{T}]_\beta$ y $[\mathsf{UT}]_\beta^\gamma$. Luego, utilizar el Teorema 2.12 para verificar el resultado.
(b) Sea $h(x) = 3 - 2x + x^2$. Calcular $[h]_\beta$ y $[\mathsf{U}(h)]_\gamma$. Luego, emplear $[\mathsf{U}]_\beta^\gamma$ de (a) y utilizar el Teorema 2.16 para verificar el resultado.

4. Para cada uno de los incisos siguientes, sea T la transformación lineal definida en el inciso correspondiente del Ejercicio 5 de la Sección 2.2. Utilizar el Teorema 2.16 para calcular:

(a) $[\mathsf{T}(A)]_\alpha$, donde $A = \begin{pmatrix} 1 & 4 \\ -1 & 6 \end{pmatrix}$.
(b) $[\mathsf{T}(f)]_\alpha$, donde $f(x) = 4 - 6x + 3x^2$.
(c) $[\mathsf{T}(A)]_\gamma$, donde $A = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$.
(d) $[\mathsf{T}(f)]_\gamma$, donde $f(x) = 6 - x + 2x^2$.

5. Completar la demostración del Teorema 2.14 y su corolario.

6. Demostrar el inciso (b) del Teorema 2.15.

7. Demostrar el Teorema 2.11.

8. Encontrar transformaciones lineales $\mathsf{U}, \mathsf{T}\colon\ \mathsf{F}^2 \to \mathsf{F}^2$ tales que $\mathsf{UT} = \mathsf{T}_0$ (la transformación nula) pero $\mathsf{TU} \neq \mathsf{T}_0$. Utilice su respuesta para encontrar matrices A y B tales que $AB = O$ pero $BA \neq O$.

9. Sea A una matriz de $m \times n$. Demostrar que A es una matriz diagonal si y sólo si $A_{ij} = \delta_{ij}A_{ii}$ para toda i y toda j.

10. Sea V un espacio vectorial y sea T: $V \to V$ lineal. Demostrar que $T^2 = T_0$ si y sólo si $R(T) \subseteq N(T)$.

11. Sean V, W y Z espacios vectoriales, y sean T: $V \to W$ y U: $W \to Z$ lineales.

(a) Si UT es uno-a-uno, demostrar que T es uno-a-uno. ¿También U debe ser uno-a-uno?
(b) Si UT es sobreyectiva, demostrar que U también lo es. ¿También T debe ser sobreyectiva?
(c) Si U y T son uno-a-uno y sobreyectivas demostrar que UT también lo es.

12. Sean A y B matrices de $n \times n$. Recordar que la traza de A, denotada por $\text{tr}(A)$, es igual a

$$\sum_{i=1}^{n} A_{ii}.$$

Demostrar que $\text{tr}(AB) = \text{tr}(BA)$ y $\text{tr}(A) = \text{tr}(A^t)$.

13. Sea V un espacio vectorial dimensionalmente finito y sea T: $V \to V$ lineal.

(a) Si $\text{rango}(T) = \text{rango}(T^2)$, demostrar que $R(T) \cap N(T) = \{0\}$. Deducir que $V = R(T) \oplus N(T)$.
(b) Demostrar que existe un entero positivo k tal que $V = R(T^k) \oplus N(T^k)$.

14.* Sea V un espacio vectorial. Determinar todas las transformaciones lineales T: $V \to V$ tales que $T = T^2$. *Sugerencia:* Nótese que $x = T(x) + (x - T(x))$ para toda x en V y demostrar que $V = \{y\colon\ T(y) = y\} \oplus N(T)$.

15. Utilizando únicamente la definición de multiplicación matricial, demostrar que la multiplicación de matrices es asociativa.

16. Para una matriz de incidencia A con la matriz asociada B definida por: $B_{ij} = 1$, si i está relacionada con j y j está relacionada con i, y de lo contrario $B_{ij} = 0$, demostrar que i pertenece a una cliqué si y sólo si $(B^3)_{ii} > 0$.

17. Utilizar el Ejercicio 16 para determinar las cliqués en las relaciones correspondientes a las siguientes matrices de incidencia.

(a) $\begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$ (b) $\begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$

18. Sea A una matriz de incidencia asociada con una relación de dominancia. Demostrar que la matriz $A + A^2$ tiene un renglón (columna) que contiene elementos positivos en todas las posiciones excepto en la diagonal principal.

19. Demostrar que la siguiente matriz A corresponde a una relación de dominancia y utilizar el Ejercicio 18 para determinar cuál(es) persona(s) domina(n) a (son dominadas por) las demás en no más de dos etapas.

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

20. Sea A una matriz de incidencia de $n \times n$ que corresponde a una relación de dominancia. Determinar el número de elementos no nulos de A.

2.4 INVERTIBILIDAD E ISOMORFISMOS

El concepto de invertibilidad se introduce muy temprano en el estudio de las funciones. Por fortuna, muchas de las propiedades intrínsecas de las funciones son compartidas por sus inversas. Por ejemplo, en cursos de cálculo aprendimos que las propiedades de continuidad o de diferenciabilidad generalmente se conservan para las funciones inversas. Veremos en esta sección (Teorema 2.19) que la inversa de una transformación lineal también es lineal. Este resultado nos ayudará de una manera importante en el estudio de las "inversas" de las matrices. Como era de esperarse de la Sección 2.3, la inversa de la transformación L_A (cuando existe) puede utilizarse para determinar las propiedades de la inversa de la matriz A.

En el resto de esta sección aplicaremos muchos de los resultados sobre invertibilidad al concepto de "isomorfismo". Veremos que los espacios vectoriales dimensionalmente finitos (sobre F) de dimensiones iguales pueden ser identificados. En breve, estas ideas serán expuestas con más precisión.

Los conceptos sobre funciones inversas que se encuentran en el apéndice *B* son, por supuesto, verdaderos para el caso de las transformaciones lineales. No obstante, repetiremos algunas de estas definiciones para emplearlas en esta sección.

Definición. *Sean* V *y* W *espacios vectoriales y sea* $T\colon V \to W$ *lineal.* T *tiene una* inversa $U\colon W \to V$ *si* $TU = I_W$ *y* $UT = I_V$. *Como en el Apéndice B, las inversas son únicas y escribiremos* $U = T^{-1}$. *Decimos que* T *es* invertible *si* T *tiene una inversa.*

Las siguientes propiedades se cumplen para funciones invertibles T y U.

1. $(TU)^{-1} = U^{-1}T^{-1}$.
2. $(T^{-1})^{-1} = T$; en particular, T^{-1} es invertible.

Utilizaremos también el hecho de que una función es invertible si y sólo si es uno-a-uno y sobreyectiva.

Ejemplo 24. Defínase $\mathsf{T}\colon \mathsf{P}_1(R) \to \mathsf{R}^2$ mediante $\mathsf{T}(a + bx) = (a, a + b)$. El lector podrá verificar directamente que $\mathsf{T}^{-1}\colon \mathsf{R}^2 \to \mathsf{P}_1(R)$ queda definida mediante $\mathsf{T}^{-1}(c, d) = c + (d - c)x$. Obsérvese que T^{-1}, también es lineal. Como lo demuestra el Teorema 2.19, esto es cierto en general.

Teorema 2.19. *Sean* V *y* W *espacios vectoriales, y sean* $\mathsf{T}\colon \mathsf{V} \to \mathsf{W}$ *lineales e invertibles, entonces* $\mathsf{T}^{-1}\colon \mathsf{W} \to \mathsf{V}$ *es lineal.*

DEMOSTRACIÓN. Sean $y_1, y_2 \in \mathsf{W}$ y $c \in F$. Como T es sobreyectiva y uno-a-uno, existen vectores únicos x_1 y x_2 tales que $\mathsf{T}(x_1) = y_1$ y $\mathsf{T}(x_2) = y_2$. Entonces $x_1 = \mathsf{T}^{-1}(y_1)$ y $x_2 = \mathsf{T}^{-1}(y_2)$, y así

$$\begin{aligned}\mathsf{T}^{-1}(cy_1 + y_2) &= \mathsf{T}^{-1}[c\mathsf{T}(x_1) + \mathsf{T}(x_2)] = \mathsf{T}^{-1}[\mathsf{T}(cx_1 + x_2)]\\ &= cx_1 + x_2 = c\mathsf{T}^{-1}(y_1) + \mathsf{T}^{-1}(y_2). \quad \blacksquare\end{aligned}$$

El Teorema siguiente se sigue inmediatamente del Teorema 2.5.

Teorema 2.20. *Sean* V *y* W *espacios vectoriales dimensionalmente finitos de dimensiones iguales, y sea* $\mathsf{T} \in \mathcal{L}(\mathsf{V}, \mathsf{W})$. *Entonces las siguientes proposiciones son equivalentes.*

(a) T *es invertible.*
(b) T *es uno-a-uno.*
(c) T *es sobreyectiva.*

Estamos ahora listos para definir la inversa de una matriz. El lector debería darse cuenta de la analogía con la inversa de una transformación lineal.

Definición. *Sea* A *una matriz de* n × n. *Entonces* A *es invertible si existe una matriz* B *de* n × n *tal que* AB = BA = I.

La matriz B es única y se llama *inversa* de A y se escribe $B = A^{-1}$. (Si C fuera otra matriz, entonces $C = CI = C(AB) = (CA)B = IB = B$.

Ejemplo 25. El lector deberá verificar que la inversa de

$$\begin{pmatrix} 5 & 7 \\ 2 & 3 \end{pmatrix} \text{ es } \begin{pmatrix} 3 & -7 \\ -2 & 5 \end{pmatrix}.$$

En la Sección 3.2 aprenderemos una técnica para calcular la inversa de una matriz. Ahora nos gustaría desarrollar una serie de resultados que relacionan inversas de matrices con inversas de transformaciones lineales.

Teorema 2.21. *Sean* V *y* W *espacios vectoriales dimensionalmente finitos con bases ordenadas* β *y* γ, *respectivamente. Sea* T: V → W *lineal. Entonces* T *es invertible si y sólo si* $[T]_\beta^\gamma$ *es invertible. Además,* $[T^{-1}]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}$.

DEMOSTRACIÓN. Supóngase que T es invertible. Como T es uno-a-uno, el Teorema 2.6 implica que $T(\beta)$ es un subconjunto independiente de W. Como T es sobreyectiva, el corolario del Teorema 2.3 implica que el subespacio generado por $L(T(\beta)) = R(T) = W$. Entonces $T(\beta)$ es una base para W con dim(V) elementos. Por lo tanto, dim(V) = dim(W). Sea $n = \dim(V)$. Entonces $[T]_\beta^\gamma$ es una matriz de $n \times n$. Ahora bien, T^{-1}: W → V satisface $TT^{-1} = I_W$ y $T^{-1}T = I_V$. De donde,

$$I_n = [I_V]_\beta = [T^{-1}T]_\beta = [T^{-1}]_\gamma^\beta [T]_\beta^\gamma.$$

Análogamente, $[T]_\beta^\gamma [T^{-1}]_\gamma^\beta = I_n$, y por lo tanto $([T]_\beta^\gamma)^{-1} = [T^{-1}]_\gamma^\beta$.

Ahora sea $A = [T]_\beta^\gamma$ invertible. Entonces existe una matriz B de $n \times n$ tal que $AB = BA = I_n$. Defínase

$$U: W \to V \quad \text{mediante} \quad U(x_j) = \sum_{i=1}^{n} B_{ij} y_i,$$

donde $\gamma = \{x_1, \ldots, x_n\}$ y $\beta = \{y_1, \ldots, y_n\}$. Entonces $[U]_\gamma^\beta = B$. Demostraremos que $U = T^{-1}$. Obsérvese que por el Teorema 2.12 $[UT]_\beta = [U]_\gamma^\beta [T]_\beta^\gamma = BA = I_n = [I_V]_\beta$. Así, $UT = I_V$, y análogamente $TU = I_W$. ■

Ejemplo 26. Para los espacios vectoriales $P_1(R)$ y R^2, selecciónese las bases $\beta = \{1, x\}$ y $\gamma = \{e_1, e_2\}$, respectivamente. Con la notación del Ejemplo 24, tenemos que

$$[T]_\beta^\gamma = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{y} \quad [T^{-1}]_\gamma^\beta = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}.$$

Se puede verificar mediante el producto matricial que cada matriz es la inversa de la otra.

Corolario 1. *Sea* V *un espacio vectorial dimensionalmente finito con una base ordenada* β, *y sea* T: V → V *lineal. Entonces* T *es invertible si y sólo si* $[T]_\beta$ *lo es. Además,* $[T^{-1}]_\beta = [T]_\beta^{-1}$.

DEMOSTRACIÓN. Ejercicio.

Corolario 2. *Sea* A *una matriz de* n × n. *Entonces* A *es invertible si y sólo si* L_A *es invertible. Además* $(L_A)^{-1} = L_{A^{-1}}$.

DEMOSTRACIÓN. Ejercicio.

La noción de invertibilidad se puede utilizar para formalizar lo que el lector ya debe haber observado, esto es, que ciertos pares de espacios vectoriales se parecen mucho entre sí, excepto por la forma de sus ele-

mentos. Por ejemplo, en el caso de $M_{2\times 2}(F)$ y F^4, si asociamos a cada matriz

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

la 4-dimensional (cuarteta) (a, b, c, d), vemos que las sumas y los productos por escalares se asocian de una manera semejante; esto es, en términos de la estructura de espacios vectoriales, estos dos espacios vectoriales pueden considerarse idénticos o "isomorfos".

Definición. *Sean* V *y* W *espacios vectoriales. Decimos que* V *es* isomorfo *a* W *si existe una transformación lineal* T: V → W *que sea invertible. Tal transformación lineal se llama* isomorfismo *de* V *en* W.

Dejamos como ejercicio la demostración del hecho de que "es isomorfo a" es una relación de equivalencia.

Ejemplo 27. Defínase T: $F^2 \to P_1(F)$ mediante $T(a_1, a_2) = a_1 + a_2x$. Es evidente que T es invertible; así F^2 es isomorfo a $P_1(F)$.

Ejemplo 28. Defínase

$$T: P_3(R) \to M_{2\times 2}(R) \qquad \text{mediante } T(f) = \begin{pmatrix} f(1) & f(2) \\ f(3) & f(4) \end{pmatrix}.$$

Se puede verificar fácilmente que T es lineal. Mediante el uso de la ecuación de interpolación de Lagrange de la Sección 1.6 puede demostrarse (compárese con el Ejercicio 20) que $T(f) = O$ solamente cuando f es el polinomio cero. Luego, T es uno-a-uno y por el Teorema 2.20 tenemos que T es invertible. Podemos concluir que $P_3(R)$ es isomorfo a $M_{2\times 2}(R)$.

En cada uno de los dos ejemplos anteriores, el lector habrá observado que los espacios vectoriales isomorfos tienen dimensiones iguales. Como lo demuestra el teorema siguiente, esto no es ninguna coincidencia.

Teorema 2.22. *Sean* V *y* W *espacios vectoriales dimensionalmente finitos (sobre el mismo campo* F*). Entonces* V *es isomorfo a* W *si y sólo si* $dim(V) = dim(W)$.

DEMOSTRACIÓN. Supóngase que V es isomorfo a W y que T: V → W es una transformación lineal uno-a-uno de V en W. Entonces como en la demostración del Teorema 2.21 tenemos que dim(V) = dim(W).

Ahora supóngase que dim(V) = dim(W) y sean $\beta = \{x_1, \ldots, x_n\}$ y $\gamma = \{y_1, \ldots, y_n\}$ bases para V y W, respectivamente. Por el Teorema 2.7 existe T: V → W tal que T es lineal y $T(x_i) = y_i$ para $i = 1, \ldots, n$. Utilizando el corolario al Teorema 2.3, tenemos que $R(T) = L(T(\beta)) = L(\gamma) = W$, por lo que T es sobreyectiva. Por el Teorema 2.5 tenemos que T también es uno-a-uno. ■

Corolario. *Si V es un espacio vectorial de dimensión* n, *entonces* V *es isomorfo a* F^n.

Hasta ahora hemos asociado transformaciones con sus representaciones matriciales. Ahora estamos en posición de demostrar que, como espacio vectorial, la colección de todas las transformaciones entre dos espacios vectoriales dados puede ser identificada con el espacio vectorial adecuado de matrices de $m \times n$.

Teorema 2.23. *Sean* V *y* W *espacios vectoriales dimensionalmente finitos de dimensiones* n *y* m, *respectivamente, y sean* β *y* γ *bases ordenadas para* V *y* W, *respectivamente. Entonces la función* Φ: $\mathcal{L}(V, W) \rightarrow M_{m \times n}(F)$, *definida por* $\Phi(T) = [T]_\beta^\gamma$ *para* $T \in \mathcal{L}(V, W)$ *es un isomorfismo.*

DEMOSTRACIÓN. El Teorema 2.9 nos permite concluir que Φ es lineal. Entonces debemos demostrar que Φ es uno-a-uno y sobreyectiva. Sean $\beta = \{x_1, \ldots, x_n\}$ y $\gamma = \{y_1, \ldots, y_m\}$.

(a) Demostraremos primero que $N(\Phi) = \{T_0\}$, donde T_0 es la transformación cero. Esto implicará que Φ es uno-a-uno. Supóngase que $\Phi(T) = O$. Entonces para cada j tenemos $T(x_j) = 0y_1 + \ldots + 0y_m = 0$. Por el corolario del Teorema 2.7 tenemos que $T = T_0$.

(b) Ahora demostraremos que Φ es sobreyectiva. Sea A una matriz de $m \times n$. De acuerdo con el Teorema 2.7 existe $T \in \mathcal{L}(V, W)$ tal que

$$T(x_j) = \sum_{i=1}^{m} A_{ij} y_i \qquad \text{para } 1 \leq j \leq n.$$

Entonces $[T]_\beta^\gamma = A$, y por lo tanto $\Phi(T) = A$, es sobreyectiva.

Corolario. *Sean* V *y* W *espacios vectoriales dimensionalmente finitos de dimensiones* m *y* n, *respectivamente. Entonces* $\mathcal{L}(V, W)$ *es dimensionalmente finito de dimensión* mn.

DEMOSTRACIÓN. La demostración se sigue de los Teoremas 2.23 y 2.22, y del hecho de que $\dim(M_{m \times n}(F)) = mn$. ■

Concluiremos esta sección con un resultado que nos permitirá ver más claramente la relación entre transformaciones lineales definidas en espacios vectoriales abstractos dimensionalmente finitos y transformaciones lineales definidas en F^n.

Principiaremos citando la transformación $x \rightarrow [x]_\beta$ discutida en la Sección 2.2.

Definición. *Sea* β *una base ordenada para un espacio vectorial* n*-dimensional* V *sobre el campo* F. *La* representación estándar de V con respecto a β *es la función* ϕ_β: $V \rightarrow F^n$ *definida por* $\phi_\beta(x) = [x]_\beta$ *para toda* $x \in V$.

Ejemplo 29. Sea $V = R^2$, $\beta = \{(1, 0), (0, 1)\}$, y $\gamma = \{(1, 2), (3, 4)\}$. Para $x = (1, -2)$ tenemos

$$\phi_\beta(x) = [x]_\beta = \begin{pmatrix} 1 \\ -2 \end{pmatrix} \quad \text{y} \quad \phi_\gamma(x) = [x]_\gamma = \begin{pmatrix} -5 \\ 2 \end{pmatrix}.$$

Ya hemos observado antes que ϕ_β es una transformación lineal. El teorema siguiente nos dice mucho más.

Teorema 2.24. *Para cualquier espacio vectorial V dimensionalmente finito con una base ordenada β, ϕ_γ es un isomorfismo.*

DEMOSTRACIÓN. Ejercicio.

Este teorema nos proporciona una prueba alternativa de que un espacio vectorial n-dimensional es isomorfo a F^n (ver el corolario del Teorema 2.22).

Estamos ahora listos para utilizar la representación estándar de un espacio vectorial junto con la representación matricial de una transformación lineal para estudiar la relación entre la transformación lineal $T: V \to W$ donde V y W son espacios vectoriales abstractos dimensionalmente finitos, y L_A: $F^n \to F^m$, donde $A = [T]_\beta^\gamma$ y β y γ son bases ordenadas cualesquiera para V y W, respectivamente.

Antes de enunciar el Teorema 2.25 consideraremos la figura 2.2. Nótese que existen dos composiciones de transformaciones lineales que mapearán a V en F^m:

1. Mapeo de V en F^n con ϕ_β y continúa esta transformación con L_A; esto nos dará la composición $L_A\phi_\beta$.
2. Mapeo de V en W con T, y continúa por ϕ_γ para obtener la composición $\phi_\gamma T$.

Estas dos composiciones están representadas por las flechas punteadas en el diagrama. El Teorema 2.25 establece que ambas composiciones dan el mismo resultado; esto es, que ambas composiciones son iguales.

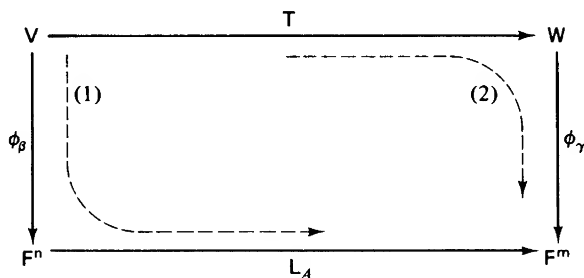


figura 2.2

Heurísticamente, el teorema establece que después de que V y W hayan sido identificados con F^n y F^m a través de ϕ_β y ϕ_γ, respectivamente, podemos "identificar" a T con L_A.

Teorema 2.25. *Sea* T: $V \to W$ *una transformación lineal de un espacio vectorial* n*-dimensional* V *sobre* F *a un espacio vectorial* m*-dimensional* W *sobre* F. *Sean* β *y* γ *bases ordenadas, respectivamente, para* V *y* W, *y sea* $A = [T]_\beta^\gamma$. *Entonces* $L_A\phi_\beta = \phi_\gamma T$.

DEMOSTRACIÓN. El teorema es esencialmente una reformulación del Teorema 2.16, porque si $x \in V$, entonces

$$(L_A\phi_\beta)(x) = L_A(\phi_\beta(x)) = L_A([x]_\beta) = A[x]_\beta = [T]_\beta^\gamma [x]_\beta$$
$$= [T(x)]_\gamma = \phi_\gamma(T(x)) = (\phi_\gamma T)(x). \quad \blacksquare$$

Ejemplo 30. Recordemos la transformación T: $P_3(R) \to P_2(R)$ definida en el Ejemplo 17. ($T(f) = f'$.) Sean $\beta = \{1, x, x^2, x^3\}$ y $\gamma = \{1, x, x^2\}$ bases ordenadas para $P_3(R)$ y $P_2(R)$, respectivamente, y sean ϕ_β: $P_3(R) \to R^4$ y ϕ_γ: $P_2(R) \to R^3$ representaciones estándar de $P_3(R)$ y $P_2(R)$ con respecto a β y γ, respectivamente. Sea $A = [T]_\beta^\gamma$; entonces

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Para ilustrar el Teorema 2.25, considérese al polinomio $p(x) = 2 + x - 3x^2 + 5x^3$. Demostraremos que $L_A\phi_\beta(p) = \phi_\gamma T(p)$.

Tenemos

$$L_A\phi_\beta(p) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ -3 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ -6 \\ 15 \end{pmatrix}.$$

Pero como

$$T(p) = p' = 1 - 6x + 15x^2,$$

tenemos que

$$\phi_\gamma T(p) = \begin{pmatrix} 1 \\ -6 \\ 15 \end{pmatrix}.$$

Así, $L_A\phi_\beta(p) = \phi_\gamma T(p)$.

Trátese de repetir este ejemplo con distintos polinomios $p(x)$.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas. Para lo siguiente, V y W serán espacios vectoriales con bases ordenadas (finitas) α y β, respectivamente, y T: V $\rightarrow$ W será lineal. A y B serán matrices.
   (a) $([\mathsf{T}]_\alpha^\beta)^{-1} = [\mathsf{T}^{-1}]_\alpha^\beta$.
   (b) T es invertible si y sólo si T es uno-a-uno y sobreyectiva.
   (c) $\mathsf{T} = \mathsf{L}_A$ donde $A = [\mathsf{T}]_\alpha^\beta$.
   (d) $\mathsf{M}_{2\times 3}(F)$ es isomorfo a F^5.
   (e) $\mathsf{P}_n(F)$ es isomorfo a $\mathsf{P}_m(F)$ si y sólo si $n = m$.
   (f) $AB = I$ implica que A y B son invertibles.
   (g) $(A^{-1})^{-1} = A$.
   (h) A es invertible si y sólo si L_A es invertible.
   (i) A debe ser cuadrada para poder tener una inversa.

2.* Sean A y B matrices invertibles de $n \times n$. Demostrar que AB es invertible y que $(AB)^{-1} = B^{-1}A^{-1}$.

3.* Sea A invertible. Demostrar que A^t es invertible y $(A^t)^{-1} = (A^{-1})^t$.

4. Demostrar que si A es invertible y $AB = O$, entonces $B = O$.

5. Si $A^2 = O$, demostrar que A no puede ser invertible.

6. Demostrar los Corolarios 1 y 2 del Teorema 2.21.

7. Sean A y B matrices de $n \times n$ tales que AB es invertible. Demostrar que A y B son invertibles. Demostrar que, en general, este resultado es falso si al menos una de las matrices no es cuadrada.

8.* Sean A y B matrices de $n \times n$ tales que $AB = I_n$. Demostrar que $A = B^{-1}$ (y por lo tanto $B = A^{-1}$). (Decimos en efecto que para matrices cuadradas una inversa unilateral es una inversa bilateral.)

9. Demostrar que la transformación definida en el Ejemplo 28 es uno-a-uno.

10. Demostrar el Teorema 2.24.

11. Sea $\sim$ tal que signifique "es isomorfo a". Demostrar que $\sim$ es una relación de equivalencia sobre la clase de espacios vectoriales sobre F tal como se define en el Apéndice A.

12. Sea

$$\mathsf{V} = \left\{ \begin{pmatrix} a & a+b \\ 0 & c \end{pmatrix} : \quad a, b, c \in F \right\}.$$

Constrúyase un isomorfismo de V a F^3.

13. Sean V y W espacios vectoriales dimensionalmente finitos y sea T: $V \to W$ un isomorfismo. Si β es una base para V, demostrar que $T(\beta)$ es una base para W.

14. Sea B una matriz invertible de $n \times n$. Defínase Φ: $M_{n \times n}(F) \to M_{n \times n}(F)$ por $\Phi(A) = B^{-1}AB$. Demostrar que Φ es un isomorfismo.

15. Sean V y W espacios vectoriales dimensionalmente finitos y sea T: $V \to W$ un isomorfismo. Sea V_0 un subespacio de V.

(a) Demostrar que $T(V_0)$ es un subespacio de W.
(b) Demostrar que $\dim(V_0) = \dim(T(V_0))$.

16. Repetir el Ejemplo 30 con el polinomio $p(x) = 1 + x + 2x^2 + x^3$.

17. Sea $V = M_{2\times 2}(R)$ el espacio vectorial de cuatro dimensiones de las matrices de 2×2 con elementos reales. Recuérdese del Ejemplo 4 que el mapeo T: $V \to V$ definido por $T(A) = A^t$ para toda $A \in V$ es una transformación lineal.

(a) Sea $\beta = \{E^{11}, E^{12}, E^{21}, E^{22}\}$ donde E^{ij} es la matriz de 2×2 que tiene el elemento i, j igual a uno y el resto de los elementos iguales a cero. Demostrar que β es una base ordenada para V.
(b) Sea $A = [T]_\beta$. Calcular A.
(c) Sea ϕ la representación estándar de V con respecto a β. Entonces $L_A\phi = \phi T$ por el Teorema 2.25. Verificar esta igualdad para la matriz

$$M = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix};$$

esto es, demostrar que $L_A\phi(M) = \phi T(M)$.

18.* Sea T: $V \to W$ una transformación lineal de un espacio vectorial n-dimensional V a un espacio m-dimensional W. Sean β y γ bases ordenadas para V y W, respectivamente. Demostrar que rango(T) = rango(L_A) y que nulidad(T) = nulidad(L_A), donde $A = [T]_\beta^\gamma$. *Sugerencia:* Utilizar el Teorema 2.25 y el Ejercicio 15.

19. Sean V y W espacios vectoriales dimensionalmente finitos con bases ordenadas $\beta = \{x_1, \ldots, x_n\}$ y $\gamma = \{y_1, \ldots, y_m\}$, respectivamente. Por el Teorema 2.7 existe una transformación lineal T_{ij}: $V \to W$ tal que

$$T_{ij}(x_k) = \begin{cases} y_i & \text{si } k = j \\ 0 & \text{si } k \neq j. \end{cases}$$

Demostrar primero que $\{T_{ij}\colon 1 \leq i \leq m,\ 1 \leq j \leq n\}$ es una base para $\mathcal{L}(V, W)$. Entonces, sea E^{ij} una matriz de $m \times n$ con 1 en el renglón

i-ésimo y la columna j-ésima y 0 en cualquier otro lado, y demostrar que $[\mathsf{T}_{ij}]_\beta^\gamma = E^{ij}$. De nuevo, por el Teorema 2.7 existe una transformación lineal Φ: $\mathcal{L}(\mathsf{V}, \mathsf{W}) \to \mathsf{M}_{m\times n}(F)$ tal que $\Phi(\mathsf{T}_{ij}) = E^{ij}$. Demostrar que Φ es un isomorfismo.

20. Sean $c_0, c_1, \ldots, c_n$ elementos distintos de un campo F. Defínase T: $\mathsf{P}_n(F) \to \mathsf{F}^{n+1}$ por $\mathsf{T}(f) = (f(c_0), \ldots, f(c_n))$. Demostrar que T es un isomorfismo. *Sugerencia:* Utilizar los polinomios de Lagrange asociados con $c_0, \ldots, c_n$.

21. Sea V el espacio vectorial de sucesiones finitas no nulas en F (definido en el Ejemplo 5 de la Sección 1.2), y sea $\mathsf{W} = \mathsf{P}(F)$. Defínase

$$\mathsf{T}: \mathsf{V} \to \mathsf{W} \qquad \text{mediante } \mathsf{T}(\sigma) = \sum_{i=0}^{n} \sigma(i)x^i,$$

donde n es el mayor de los enteros con una imagen no nula. Demostrar que T es un isomorfismo.

2.5 LA MATRIZ DE CAMBIO DE COORDENADAS

En muchas áreas de las matemáticas se utiliza a menudo un cambio de variable para simplificar la apariencia de una expresión. Por ejemplo, en cálculo puede encontrarse fácilmente una antiderivada de $2xe^{x^2}$ haciendo el cambio de variable $u = x^2$. La expresión que resulta tiene una forma tan sencilla que la antiderivada se reconoce fácilmente:

$$\int 2xe^{x^2}\,dx = \int e^u\,du = e^u = e^{x^2}.$$

De la misma forma, en geometría plana puede emplearse el cambio de variable

$$x = \frac{\sqrt{5}}{5}x' + \frac{2\sqrt{5}}{5}y' \quad \text{y} \quad y = \frac{-2\sqrt{5}}{5}x' + \frac{\sqrt{5}}{5}y'$$

para transformar la ecuación $2x^2 - 4xy + 5y^2 = 1$ en la ecuación más sencilla $6(x')^2 + (y')^2 = 1$ de cuya forma se reconoce fácilmente que se trata de la ecuación de una elipse. (Veremos en la Sección 7.7 cómo se determinó este cambio de variable.) Geométricamente, el cambio de variable

$$\begin{pmatrix} x \\ y \end{pmatrix} \to \begin{pmatrix} x' \\ y' \end{pmatrix}$$

equivale a una rotación de los ejes coordenados de manera que los ejes x y y coincidan con los ejes x' y y', respectivamente, donde

$$x' = \frac{\sqrt{5}}{5}x - \frac{2\sqrt{5}}{5}y \quad \text{y} \quad y' = \frac{2\sqrt{5}}{5}x + \frac{\sqrt{5}}{5}y.$$

(Los valores de x' y y' se encontraron al resolver el sistema

$$\begin{cases} \dfrac{\sqrt{5}}{5}x' + \dfrac{2\sqrt{5}}{5}y' = x \\ \dfrac{-2\sqrt{5}}{2}x' + \dfrac{\sqrt{5}}{5}y' = y \end{cases}$$

para x' y y' en términos de x y y.) Los vectores obtenidos al someter los vectores de la base

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ y } \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

del sistema coordenado x, y a esta rotación son

$$\begin{pmatrix} \dfrac{\sqrt{5}}{5} \\ \dfrac{2\sqrt{5}}{5} \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} \dfrac{-2\sqrt{5}}{5} \\ \dfrac{\sqrt{5}}{5} \end{pmatrix},$$

respectivamente. Estos vectores son vectores unitarios que se alojan en el eje x' y el eje y', respectivamente, y por lo tanto forman una nueva base

$$\beta' = \left\{ \begin{pmatrix} \dfrac{\sqrt{5}}{5} \\ \dfrac{2\sqrt{5}}{5} \end{pmatrix}, \begin{pmatrix} \dfrac{-2\sqrt{5}}{5} \\ \dfrac{\sqrt{5}}{5} \end{pmatrix} \right\}$$

para R^2.

Surge una interrogante lógica: ¿Cómo es posible transformar vectores coordenados con respecto a una base, a vectores coordenados con respecto a la otra? La respuesta está dada a través de la relación

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \dfrac{\sqrt{5}}{5} & \dfrac{-2\sqrt{5}}{5} \\ \dfrac{2\sqrt{5}}{5} & \dfrac{\sqrt{5}}{5} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Nótese que la matriz

$$Q = \begin{pmatrix} \dfrac{\sqrt{5}}{5} & \dfrac{-2\sqrt{5}}{5} \\ \dfrac{2\sqrt{5}}{5} & \dfrac{\sqrt{5}}{5} \end{pmatrix}$$

es igual a $[I]_{\beta'}^{\beta}$, donde I es la transformación identidad en R^2. Entonces por el Teorema 2.16 $[v]_\beta = Q[v]_{\beta'}$ para toda $v \in R^2$. Un resultado similar es cierto en general.

Teorema 2.26. *Sean β y β' dos bases ordenadas para un espacio vectorial dimensionalmente finito V y sea $Q = [I_V]_{\beta'}^{\beta}$. Entonces*

(a) Q *es invertible.*
(b) *Para toda* $v \in V$, $[v]_\beta = Q[v]_{\beta'}$.

DEMOSTRACIÓN.

(a) Como I_V es invertible, Q también lo es por el Teorema 2.21.
(b) Para toda $v \in V$,

$$[v]_\beta = [I_V(v)]_\beta = [I_V]_{\beta'}^{\beta}[v]_{\beta'} = Q[v]_{\beta'}$$

por el Teorema 2.16. ■

La matriz Q definida en el Teorema 2.26 se llama *matriz de cambio de coordenadas*. A causa del inciso (b) del teorema decimos que *Q transforma las coordenadas de* β' *en coordenadas de* β. Obsérvese que si $\beta = \{x_1, x_2, \ldots, x_n\}$ y $\beta' = \{x'_1, x'_2, \ldots, x'_n\}$, entonces

$$x'_j = \sum_{i=1}^{n} Q_{ij} x_i$$

para $j = 1, 2, \ldots\ n$; esto es, la columna j-ésima de Q es $[x'_j]_\beta$.

Ejemplo 31. Sea $V = R^2$, $\beta = \{(1, 1), (1, -1)\}$, y $\beta' = \{(2, 4), (3, 1)\}$. Como $(2, 4) = 3(1, 1) - 1(1, -1)$ y $(3, 1) = 2(1, 1) + 1(1, -1)$, la matriz que transforma las coordenadas de β' en coordenadas de β es

$$Q = \begin{pmatrix} 3 & 2 \\ -1 & 1 \end{pmatrix}.$$

Así, por ejemplo,

$$[(2, 4)]_\beta = Q[(2, 4)]_{\beta'} = Q\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}.$$

Supóngase ahora que $T\colon V \to W$ es una transformación lineal entre espacios vectoriales dimensionalmente finitos y que β y β' son bases ordenadas para V y γ y γ' son bases ordenadas para W. Entonces T se puede representar por matrices relativas a β y γ y relativas a β' y γ'. ¿Cuál es la relación entre las matrices $[T]_\beta^\gamma$ y $[T]_{\beta'}^{\gamma'}$? La respuesta se ve claramente de las ecuaciones $[T(v)]_\gamma = [T]_\beta^\gamma[v]_\beta$ y $[T(v)]_{\gamma'} = [T]_{\beta'}^{\gamma'}[v]_{\beta'}$ dadas por el Teorema 2.16, porque si Q y P son matrices de cambio de coordenadas que transforman coordenadas de β' en coordenadas de β y coordenadas de γ' en coordenadas γ, respectivamente, entonces de estas ecuaciones es evidente que hay dos métodos para obtener $[T(v)]_\gamma$ a partir de $[v]_{\beta'}$, como se ilustra en la figura 2.3.

Como $[T]_\beta^\gamma Q[v]_{\beta'} = P[T]_{\beta'}^{\gamma'}[v]_{\beta'}$ para toda $v \in V$, el Teorema 2.17(b) implica que $[T]_\beta^\gamma Q = P[T]_{\beta'}^{\gamma'}$. Al ser P invertible (Teorema 2.26), esto da la respuesta a la pregunta anterior.

Teorema 2.27. *Sea* T: V → W *una transformación lineal de un espacio vectorial* V *dimensionalmente finito a un espacio vectorial* W *dimensionalmente finito, y sean* β *y* β' *bases ordenadas para* V *y* γ *y* γ' *bases ordenadas para* W. *Entonces* $[T]_{\beta'}^{\gamma'} = P^{-1}[T]_{\beta}^{\gamma}Q$, *donde* Q *es la matriz que transforma las coordenadas de* β' *en coordenadas de* β *y* P *es la matriz que transforma las coordenadas de* γ' *en coordenadas de* γ.

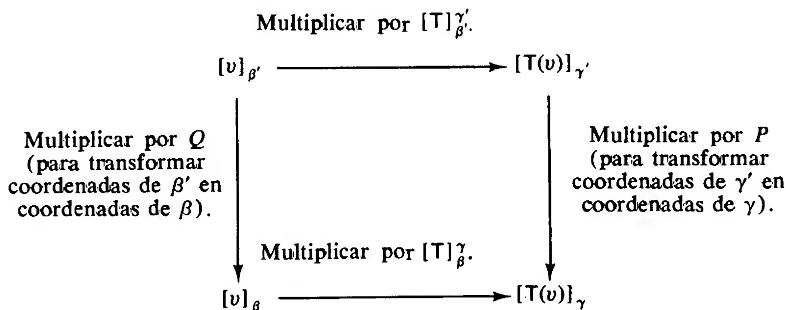


figura 2.3

Ejemplo 32. Sean V = R^3, W = R^2 y T: V → W definida mediante

$$T\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 2a_1 + a_2 \\ a_1 + a_2 - a_3 \end{pmatrix}.$$

Sean β y γ las bases ordenadas estándar para R^3 y R^2, respectivamente, y sean

$$\beta' = \left\{\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -2 \end{pmatrix}\right\} \quad \text{y} \quad \gamma' = \left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}\right\}.$$

(Obsérvese que β' y γ' son bases ordenadas para R^3 y R^2, respectivamente.) Verificaremos el Teorema 2.27. Tenemos que

$$[T]_{\beta}^{\gamma} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 1 & -1 \end{pmatrix} \quad \text{y} \quad [T]_{\beta'}^{\gamma'} = \begin{pmatrix} 3 & -1 & 1 \\ 1 & 0 & -1 \end{pmatrix},$$

y algunos cálculos fáciles muestran que las matrices de cambio de coordenadas que transforman coordenadas de β' en coordenadas de β y coordenadas de γ' en coordenadas de γ son

$$Q = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \quad \text{y} \quad P = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

respectivamente. En la Sección 3.2 se presentará un método para calcular P^{-1}; mientras tanto, verifíquese que

$$P^{-1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}.$$

Una multiplicación fácil muestra que $[T]_{\beta'}^{\gamma'} = P^{-1}[T]_{\beta}^{\gamma}Q$.

Un caso particular importante del resultado anterior se tiene cuando $V = W$. Es esta la situación que nos ocupará primordialmente en una gran parte del resto del libro. En este caso el teorema toma la siguiente forma.

Corolario. *Sea* $T\colon V \to V$ *una transformación lineal en un espacio vectorial dimensionalmente finito* V *que tiene bases ordenadas* β *y* β'. *Entonces* $[T]_{\beta'} = Q^{-1}[T]_{\beta}Q$, *donde* Q *es la matriz que transforma coordenadas de* β' *en coordenadas de* β.

La relación entre las matrices $[T]_{\beta'}$ y $[T]_{\beta}$ del corolario anterior serán temas de estudio en los Capítulos 5 y 6. Sin embargo, introduciremos ahora el nombre de esta relación.

Definición. *Sean* A *y* B *matrices de* $n \times n$ *con elementos del campo* F. *Decimos que* B *es similar a* A *si existe una matriz invertible* $Q \in M_{n \times n}(F)$ *tal que* $B = Q^{-1}AQ$.

Obsérvese que la relación de similaridad es una relación de equivalencia. (Ver el Ejercicio 7.)

Nótese también que en esta terminología el corolario anterior puede enunciarse de la manera siguiente: Si $T\colon V \to V$ es una transformación lineal en un espacio vectorial dimensionalmente finito V, si β y β' son bases ordenadas cualesquiera para V, entonces $[T]_{\beta'}$ es similar a $[T]_{\beta}$.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

   (a) Si Q es la matriz de cambio de coordenadas que transforma las coordenadas de β' en coordenadas de β, donde $\beta' = \{x'_1, \ldots, x'_n\}$ y $\beta = \{x_1, \ldots, x_n\}$ son bases ordenadas para un espacio vectorial, entonces la columna j-ésima de Q es $[x_j]_{\beta'}$.
   (b) Toda matriz de cambio de coordenadas es invertible.
   (c) Sea $T\colon V \to W$ una transformación lineal de un espacio vectorial dimensionalmente finito V a un espacio vectorial dimensionalmente finito W, y sean β y β' bases ordenadas para V y γ y γ' bases ordenadas para W. Entonces $[T]_{\beta'}^{\gamma'} = P[T]_{\beta}^{\gamma}Q$, donde Q y P son las matrices de cambio de coordenadas que transforman coordenadas de β' en coor-

denadas de β y coordenadas de γ en coordenadas de γ', respectivamente.

(d) Las matrices A, $B \in \mathsf{M}_{n\times n}(F)$ se llaman similares si $B = Q^tAQ$ para alguna $Q \in \mathsf{M}_{n\times n}(F)$.

(e) Sea T: V → V una transformación lineal en un espacio vectorial dimensionalmente finito V. Entonces para cualquier par de bases ordenadas β y γ para V, $[\mathsf{T}]_\beta$ es similar a $[\mathsf{T}]_\gamma$.

2. Para cada uno de los siguientes pares de bases ordenadas β y β' para R^2, encontrar la matriz de cambio de coordenadas que transforma coordenadas de β' en coordenadas de β.

(a) $\beta = \{e_1, e_2\}$ y $\beta' = \{(a_1, a_2), (b_1, b_2)\}$
(b) $\beta = \{(-1, 3), (2, -1)\}$ y $\beta' = \{(0, 10), (5, 0)\}$
(c) $\beta = \{(2, 5), (-1, -3)\}$ y $\beta' = \{e_1, e_2\}$
(d) $\beta = \{(-4, 3), (2, -1)\}$ y $\beta' = \{(2, 1), (-4, 1)\}$

3. Para cada uno de los siguientes pares de bases ordenadas β y β' para $\mathsf{P}_2(R)$, encontrar la matriz de cambio de coordenadas que transforma las coordenadas de β' en coordenadas de β.

(a) $\beta = \{x^2, x, 1\}$ y
$\beta' = \{a_2x^2 + a_1x + a_0, b_2x^2 + b_1x + b_0, c_2x^2 + c_1x + c_0\}$
(b) $\beta = \{1, x, x^2\}$ y
$\beta' = \{a_2x^2 + a_1x + a_0, b_2x^2 + b_1x + b_0, c_2x^2 + c_1x + c_0\}$
(c) $\beta = \{2x^2 - x, 3x^2 + 1, x^2\}$ y $\beta' = \{1, x, x^2\}$
(d) $\beta = \{x^2 - x + 1, x + 1, x^2 + 1\}$ y $\beta' = \{x^2 + x + 4, 4x^2 - 3x + 2, 2x^2 + 3\}$
(e) $\beta = \{x^2 - x, x^2 + 1, x - 1\}$ y $\beta' = \{5x^2 - 2x - 3, -2x^2 + 5x + 5, 2x^2 - x - 3\}$
(f) $\beta = \{2x^2 - x + 1, x^2 + 3x - 2, -x^2 + 2x + 1\}$ y $\beta' = \{9x - 9, x^2 + 21x - 2, 3x^2 + 5x + 2\}$

4. Sea V = R^2, W = R^3 y T: V → W esté definida mediante

$$\mathsf{T}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 3a_1 - a_2 \\ 2a_1 + 4a_2 \\ -a_1 + a_2 \end{pmatrix}.$$

Sean β y γ bases ordenadas estándar para R^2 y R^3, respectivamente, y sean

$$\beta' = \left\{\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\} \quad \text{y} \quad \gamma' = \left\{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\right\}.$$

(a) Calcular $A = [T]_\beta^\gamma$ y $B = [T]_{\beta'}^{\gamma'}$.
(b) Calcular Q, la matriz de cambio de coordenadas que transforma las coordenadas de β' en coordenadas de β, y P, la matriz de cambio de coordenadas que transforma coordenadas de γ' en coordenadas de γ.
(c) Verificar que $B = P^{-1}AQ$, considerando que

$$P^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix}.$$

5. Sea T: $P_1(R) \to P_1(R)$ definida por $T(p) = p'$, la derivada de $p \in P_1(R)$. Sean $\beta = \{1, x\}$ y $\beta' = \{1 + x, 1 - x\}$.

(a) Encontrar la matriz de cambio de coordenadas Q que transforma las coordenadas de β' en coordenadas de β.
(b) Encontrar Q^{-1}. (Ver el Ejemplo 32.)
(c) Calcular $A = [T]_\beta$ y $B = [T]_{\beta'}$ y verificar que $B = Q^{-1}AQ$.

6. Demostrar el corolario del Teorema 2.27.

7. Recordando la definición de una relación de equivalencia dada en el Apéndice A, demostrar que la relación "es similar a" es una relación de equivalencia sobre $M_{n\times n}(F)$.

8. Demostrar que si A y B son matrices semejantes de $n \times n$, entonces $\text{tr}(A) = \text{tr}(B)$. *Sugerencia:* Utilizar el Ejercicio 12 de la Sección 2.3.

9. Sea V un espacio vectorial dimensionalmente finito con bases ordenadas α, β y γ.

(a) Demostrar que si Q y R son las matrices de cambio de coordenadas que transforman coordenadas de α en coordenadas de β y coordenadas de β en coordenadas de γ, respectivamente, entonces RQ es la matriz de cambio de coordenadas que transforma coordenadas de α en coordenadas de γ.
(b) Demostrar que si Q transforma coordenadas de α en coordenadas de β, entonces Q^{-1} transforma coordenadas de β en coordenadas de α.

10. Sea A una matriz de $m \times n$ con elementos de un campo F, y sean β y γ bases ordenadas para F^n y F^m, respectivamente. Sea $B = [L_A]_\beta^\gamma$. Demostrar que $B = P^{-1}AQ$ donde P es la matriz de $m \times m$ con la columna j-ésima igual al vector j-ésimo en β y Q es la matriz de $n \times n$ con la columna j-ésima igual al vector j-ésimo en γ.

11.* Sea V un espacio vectorial dimensionalmente finito sobre un campo F y sea $\beta = \{x_1, \ldots, x_n\}$ una base ordenada para V. Sea Q una matriz invertible de $n \times n$ con elementos de F. Defínase

$$x'_j = \sum_{i=1}^{n} Q_{ij} x_i \qquad \text{para } 1 \leq j \leq n,$$

y hágase $\beta' = \{x'_1, \ldots, x'_n\}$. Demostrar que β' es una base para V y por lo tanto Q es la matriz de cambio de coordenadas que transforma coordenadas de β' en coordenadas de β.

12.* Demostrar la recíproca de Teorema 2.27: Si A y B son ambas matrices de $m \times n$ sobre un campo F y existen matrices invertibles P y Q de $m \times m$ y $n \times n$, respectivamente, tales que $B = PAQ$, entonces existen un espacio vectorial n-dimensional V y un espacio vectorial m-dimensional W (ambos sobre F), bases ordenadas β y β' para V y γ y γ' para W y una transformación lineal T: V $\rightarrow$ W tal que

$$A = [\mathsf{T}]_\beta^\gamma \quad \text{y} \quad B = [\mathsf{T}]_{\beta'}^{\gamma'}.$$

Sugerencias: Sean V $=$ F^n, W $=$ F^m, T $=$ L$_A$ y β y γ bases ordenadas estándar para F^n y F^m, respectivamente. Sea β' una base ordenada para V obtenida a partir de β a través de Q (de acuerdo con la definición dada en la página 106 y justificada por el Ejercicio 11), y sea γ' la base para W obtenida de γ a través de P^{-1}.

2.6* ESPACIOS DUALES

En esta sección nos interesaremos exclusivamente en las transformaciones lineales de un espacio vectorial V en su campo de escalares F, que a su vez es un espacio vectorial de dimensión 1 sobre F. Tal transformación lineal se llama *funcional lineal en* V. En el cálculo, la integral definida nos proporciona uno de los ejemplos más importantes en matemáticas de una funcional lineal. (Ver Ejemplo 33.) Utilizaremos generalmente las letras f, g, h, ... para denotar a las funcionales lineales.

Ejemplo 33. Sea V el espacio vectorial de funciones continuas complejas (o reales) sobre el intervalo $[a, b]$. La función f: V $\rightarrow C$ (o R) definida por

$$\mathsf{f}(x) = \int_a^b x(t)\,dt$$

es una funcional lineal en V. Si el intervalo es $[0, 2\pi]$ y n es un entero, la función definida por

$$\mathsf{h}_n(x) = \frac{1}{2\pi} \int_0^{2\pi} x(t) e^{-int}\,dt$$

también es una funcional lineal. En los textos de análisis al escalar $h_n(x)$ se le llama el n-ésimo *coeficiente de Fourier de x.*

Ejemplo 34. Sea $V = M_{m \times n}(F)$ y defínase $f: V \to F$ por $f(A) = tr(A)$, la traza de A. Por el Ejercicio 6 de la Sección 1.3, tenemos que f es una funcional lineal.

Ejemplo 35. Sea V un espacio vectorial dimensionalmente finito con la base ordenada $\beta = \{x_1, x_2, \ldots, x_n\}$. Para cada $i = 1, \ldots, n$, defínase $f_i(x) = a_i$, donde

$$[x]_\beta = \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{pmatrix}$$

es el vector de coordenadas de x relativo a β. Entonces f_i es una funcional lineal en V llamada la i-*ésima función coordenada con respecto a la base* β. Nótese que $f_i(x_j) = \delta_{ij}$. Estas funcionales lineales jugarán un papel muy importante en la teoría de los espacios duales. (Ver el Teorema 2.28.)

Definición. *Para un espacio vectorial* V *sobre* F, *definimos al* espacio dual de V *como el espacio vectorial* $\mathcal{L}(V, F)$, *denotado por* V^*.

Por tanto, V^* es el espacio vectorial que consta de todas las funcionales lineales en V con las operaciones de suma y de multiplicación por escalares tal como se definieron en la Sección 2.2. Nótese que si V es dimensionalmente finito, entonces $\dim(V^*) = \dim(\mathcal{L}(V, F)) = \dim(V) \cdot \dim(F) = \dim(V)$. Por lo tanto, por el Teorema 2.22, V y V^* son isomorfos. También podemos definir el *doble dual* V^{**} de V como el dual de V^*. Demostraremos, de hecho, que existe una identificación natural de V y V^{**}.

Teorema 2.28. *Supóngase que* V *es un espacio vectorial dimensionalmente finito con la base ordenada* $\beta = \{x_1, \ldots, x_n\}$. *Sean* $f_i (1 \leq i \leq n)$ *las funciones coordenadas con respecto a* β *tal como se definieron anteriormente y sea* $\beta^* = \{f_1, \ldots, f_n\}$. *Entonces* β^* *es una base ordenada para* V^*, *y para cualquier* $f \in V^*$ *tenemos que*

$$f = \sum_{i=1}^{n} f(x_i) f_i.$$

Llamamos a β^* *la base dual de* β.

DEMOSTRACIÓN. Sea $f \in V^*$. Como $\dim(V^*) = n$, sólo necesitamos probar que

$$f = \sum_{i=1}^{n} f(x_i)f_i,$$

porque entonces β^* generará a V^*. Sea

$$g = \sum_{i=1}^{n} f(x_i)f_i.$$

Para $1 \leq j \leq n$, tenemos que

$$g(x_j) = \left(\sum_{i=1}^{n} f(x_i)f_i\right)(x_j) = \sum_{i=1}^{n} f(x_i)f_i(x_j)$$
$$= \sum_{i=1}^{n} f(x_i)\delta_{ij} = f(x_j).$$

Por lo tanto, por el corolario del Teorema 2.7, $g = f$, por lo que el teorema queda demostrado. ■

Ejemplo 36. Sea $\beta = \{(2, 1), (3, 1)\}$ una base ordenada para R^2. Determinaremos explícitamente la base dual $\beta^* = \{f_1, f_2\}$ de β. Necesitamos considerar las ecuaciones:

$$1 = f_1(2, 1) = f_1(2e_1 + e_2) = 2f_1(e_1) + f_1(e_2)$$
$$0 = f_1(3, 1) = f_1(3e_1 + e_2) = 3f_1(e_1) + f_1(e_2).$$

Resolviéndolas tenemos que $f_1(e_1) = -1$ y $f_1(e_2) = 3$ es decir que $f_1(x, y) = -x + 3y$. De manera semejante puede probarse que $f_2(x, y) = x - 2y$.

Ahora supondremos que V y W son espacios vectoriales dimensionalmente finitos sobre F con bases ordenadas β y γ, respectivamente. En la Sección 2.4 demostramos que existe una correspondencia uno-a-uno entre las transformaciones lineales $T\colon V \to W$ y las matrices de $m \times n$ (sobre F) por medio de la correspondencia $T \leftrightarrow [T]_\beta^\gamma$. Para una matriz de la forma $A = [T]_\beta^\gamma$, la pregunta es si existe o no una transformación lineal U asociada con T de alguna manera natural tal que U pueda representarse en alguna base como A^t. Por supuesto, si $m \neq n$, sería imposible para U ser una transformación lineal de V en W. Resolveremos esta pregunta aplicando lo que ya hemos aprendido sobre espacios duales.

Teorema 2.29. *Sean V y W espacios vectoriales dimensionalmente finitos sobre F con bases ordenadas β y γ, respectivamente. Para cualquier transformación lineal $T\colon V \to W$, el mapeo $T^t\colon W^* \to V^*$ definido por $T^t(g) = g \circ T$ para toda $g \in W^*$ es una transformación lineal con la propiedad de que $[T^t]_{\gamma^*}^{\beta^*} = ([T]_\beta^\gamma)^t$.*

DEMOSTRACIÓN. Para $g \in W^*$, es evidente que $T^t(g) = g \circ T$ es una función lineal en V y por lo tanto es un elemento de V^*. Así, T^t mapea a W^* en V^*. Dejaremos al lector la demostración de que T^t es lineal.

Para completar la demostración sean $\beta = \{x_1, \ldots, x_n\}$ y $\gamma = \{y_1, \ldots, y_m\}$ con bases duales $\beta^* = \{f_1, \ldots, f_n\}$ y $\gamma^* = \{g_1, \ldots, g_m\}$, respectivamente. Por conveniencia, sean $A = [T]_\beta^\gamma$ y $B = [T^t]_{\gamma^*}^{\beta^*}$. Entonces

$$T(x_i) = \sum_{k=1}^{m} A_{ki} y_k \qquad \text{para } 1 \leq i \leq n,$$

y

$$T^t(g_j) = \sum_{i=1}^{n} B_{ij} f_i \qquad \text{para } 1 \leq j \leq m.$$

Debemos demostrar que $B = A^t$. El Teorema 2.28 muestra que

$$T^t(g_j) = g_j \circ T = \sum_{i=1}^{n} (g_j \circ T)(x_i) f_i,$$

y entonces

$$B_{ij} = (g_j \circ T)(x_i) = g_j(T(x_i)) = g_j\left(\sum_{k=1}^{m} A_{ki} y_k\right)$$

$$= \sum_{k=1}^{m} A_{ki} g_j(y_k) = \sum_{k=1}^{m} A_{ki} \delta_{jk} = A_{ji} = (A^t)_{ij}.$$

Por lo tanto $B = A^t$. ■

La transformación lineal T^t definida en el Teorema 2.29 se llama *transpuesta de* T. Es evidente que T^t es la única transformación lineal U tal que $[U]_{\gamma^*}^{\beta^*} = ([T]_\beta^\gamma)^t$.

Ahora nos ocuparemos de la demostración de que cualquier espacio vectorial dimensionalmente finito V puede ser identificado de una manera muy natural con su doble dual V^{**}. Produciremos, de hecho, un isomorfismo entre V y V^{**} que no dependerá de ninguna selección de bases para los dos espacios vectoriales.

Para un vector $x \in V$ definimos $\hat{x}$: $V^* \to F$ por $\hat{x}(f) = f(x)$ para toda $f \in V^*$. Es fácil verificar que $\hat{x}$ es una funcional lineal en V^* y entonces $\hat{x} \in V^{**}$. La correspondencia $x \leftrightarrow \hat{x}$ nos permitirá definir el isomorfismo deseado entre V y V^{**}.

Lema. *Sea* V *un espacio vectorial dimensionalmente finito y sea* $x \in V$. *Si* $\hat{x}(f) = 0$ *para toda* $f \in V^*$, *entonces* $x = 0$.

DEMOSTRACIÓN. Si $x \neq 0$ entonces podemos tomar una base ordenada $\beta = \{x_1, \ldots, x_n\}$ para V tal que $x_1 = x$. Sea $\{f_1, \ldots, f_n\}$ la base dual de β. Entonces se tiene que $f_1(x_1) = 1 \neq 0$ lo que es una contradicción.

Teorema 2.30. *Sea V un espacio vectorial dimensionalmente finito y sea ψ: $V \rightarrow V^{**}$ que esté definida por $\psi(x) = \hat{x}$. Entonces ψ es un isomorfismo.*

DEMOSTRACIÓN.

(a) ψ es lineal: Sea $x, y \in V$ y $a \in F$. Para $f \in V^*$ tenemos que

$$\begin{aligned}\psi(x + ay)(f) &= f(x + ay) = f(x) + af(y) = \hat{x}(f) + a\hat{y}(f) \\ &= (\hat{x} + a\hat{y})(f).\end{aligned}$$

Por lo tanto

$$\psi(x + ay) = \hat{x} + a\hat{y} = \psi(x) + a\psi(y).$$

(b) ψ es uno-a-uno: Supóngase que $\psi(x)$ es la funcional cero en V^* para alguna $x \in V$. Entonces $\hat{x}(f) = 0$ para toda $f \in V^*$. Por el lema anterior concluimos que $x = 0$.

(c) ψ es un isomorfismo: Esto se deduce de (b) y del hecho de que $\dim(V) = \dim(V^{**})$. ■

Corolario. *Sea V un espacio vectorial dimensionalmente finito con un espacio dual V^*. Entonces, toda base ordenada de V^* es la base dual de alguna base de V.*

DEMOSTRACIÓN. Sea $\{f_1, \ldots, f_n\}$ una base ordenada de V^*. Podemos combinar los Teoremas 2.28 y 2.30 para concluir que para esta base de V^* existe una base dual $\{\hat{x}_1, \ldots, \hat{x}_n\}$ en V^{**}, esto es $\delta_{ij} = \hat{x}_i(f_j) = f_j(x_i)$. Por tanto $\{f_1, \ldots, f_n\}$ es la base dual de $\{x_1, \ldots, x_n\}$. ■

Aun cuando muchas de las ideas de esta sección se pueden extender para el caso donde V no es dimensionalmente finito, por ejemplo la existencia de un espacio dual, únicamente un espacio vectorial dimensionalmente finito es isomorfo a su doble dual a través del mapeo $x \longrightarrow \hat{x}$. De hecho, para espacios vectoriales dimensionalmente infinitos, V y V^* nunca son isomorfos.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas. Supóngase que todos los espacios vectoriales son dimensionalmente finitos.

   (a) Toda transformación lineal es una funcional lineal.
   (b) Una funcional lineal definida en un campo puede ser representada como una matriz de 1×1.
   (c) Todo espacio vectorial es isomorfo a su espacio dual.
   (d) Todo espacio vectorial es el dual de algún otro espacio vectorial.
   (e) Si T es un isomorfismo de V en V^* y β es una base ordenada finita de V, entonces $T(\beta) = \beta^*$.

(f) Si T es una transformación lineal de V en W, entonces el dominio de $(T^t)^t$ es V**.
(g) Si V es isomorfo a W, entonces V* es isomorfo a W*.
(h) La derivada de una función puede considerarse como una funcional lineal en el espacio vectorial de funciones diferenciables.

2. Para las siguientes funciones en un espacio vectorial V, determinar cuáles son funcionales lineales.

(a) $V = P(R)$; $f(p) = 2p'(0) + p''(1)$, donde $'$ significa diferenciación
(b) $V = R^2$; $f(x, y) = (2x, 4y)$
(c) $V = M_{2\times 2}(F)$; $f(A) = \mathrm{tr}(A)$
(d) $V = R^3$; $f(x, y, z) = x^2 + y^2 + z^2$
(e) $V = P(R)$; $f(p) = \int_0^1 p(t)dt$
(f) $V = M_{2\times 2}(R)$; $f(A) = A_{11}$

3. Como en el Ejemplo 36, para todos los espacios vectoriales V y bases β que aparecen a continuación, encontrar la base dual β^* para V*.

(a) $V = R^3$; $\beta = \{(1, 0, 1), (1, 2, 1), (0, 0, 1)\}$
(b) $V = P_2(R)$; $\beta = \{1, x, x^2\}$

4. Sea $V = R^3$ y defínase $f_1, f_2, f_3 \in V^*$ mediante $f_1(x, y, z) = x - 2y$, $f_2(x, y, z) = x + y + z$, y $f_3(x, y, z) = y - 3z$. Demostrar que $\{f_1, f_2, f_3\}$ es una base para V* y luego encontrar una base para V para la cual sea el dual.

5. Sea $V = P_1(R)$ y para $p \in V$ defínase $f_1, f_2 \in V^*$ mediante

$$f_1(p) = \int_0^1 p(t)dt$$

y

$$f_2(p) = \int_0^2 p(t)dt.$$

Demostrar que $\{f_1, f_2\}$ es una base para V* y encontrar una base para V para la cual sea el dual.

6. Defínase $f \in (R^2)^*$ mediante $f(x, y) = 2x + y$ y $T: R^2 \to R^2$ mediante $T(x, y) = (3x + 2y, x)$.

(a) Calcular $T^t(f)$.
(b) Calcular $[T^t]_{\beta^*}$, donde β es la base ordenada estándar para R^2 y $\beta^* = \{f_1, f_2\}$ encontrando escalares a, b, c y d tales que $T^t(f_1) = af_1 + bf_2$ y $T^t(f_2) = cf_1 + df_2$.
(c) Calcular $[T]_\beta$ y $[T]_\beta^t$ y comparar los resultados con los del inciso (b).

7. Sean $V = P_1(R)$ y $W = R^2$ con sus respectivas bases ordenadas $\beta = \{1, x\}$ y $\gamma = \{e_1, e_2\}$. Defínase $T: V \to W$ mediante $T(p) = (p(0) - 2p(1), p(0) + p'(0))$, donde p' es la derivada de p.

(a) Si $f \in W^*$ está definida por

$$f(a, b) = a - 2b,$$

calcular $T^t(f)$.

(b) Calcular $[T^t]_{\gamma*}^{\beta*}$ sin recurrir al Teorema 2.29.

(c) Calcular $[T]_\beta^\gamma$ y su transpuesta y comparar el resultado con el del inciso (b).

8. Demostrar que todo plano que pasa por el origen en R^3 puede ser identificado con el espacio nulo de un elemento en $(R^3)^*$. Enunciar un resultado semejante en R^2.

9. Sea T una función de F^n en F^m. Demostrar que T es lineal si y sólo si existen $f_1, \ldots, f_m \in (F^n)^*$ tales que $T(x) = (f_1(x), \ldots, f_m(x))$ para toda $x \in F^n$. *Sugerencia:* Si T es lineal, defínase $f_i(x) = (g_i \circ T)(x)$ para $x \in F^n$; es decir, $f_i = T^t(g_i)$ para $1 \leq i \leq m$, donde $\{g_1, \ldots, g_m\}$ es la base dual de la base ordenada estándar de F^m.

10. Sea $V = P_n(F)$ y sean $c_0, \ldots, c_n$, escalares distintos en F.

(a) Para $0 \leq i \leq n$ defínase $f_i \in V^*$ mediante $f_i(p) = p(c_i)$. Demostrar que $\{f_0, \ldots, f_n\}$ es una base de V^*. *Sugerencia:* Aplicar cualquier combinación lineal de este conjunto que iguale la transformación cero con $p(t) = (t - c_1)(t - c_2) \cdots (t - c_n)$ y deducir que el primer coeficiente es cero.

(b) Utilizar el corolario del Teorema 2.30 y el inciso (a) para demostrar que existen polinomios únicos $p_0, \ldots, p_n$ tales que $p_i(c_j) = \delta_{ij}$ para $0 \leq i \leq n$. Estos polinomios son los polinomios de Lagrange definidos en la Sección 1.6.

(c) Para escalares cualesquiera $a_0, \ldots, a_n$ (no necesariamente distintos), dedúzcase que existe un polinomio único q de grado n tal que $q(c_i) = a_i$ para $0 \leq i \leq n$. De hecho

$$q = \sum_{i=0}^{n} a_i p_i.$$

(d) Deducir la fórmula de interpolación de Lagrange:

$$p = \sum_{i=0}^{n} p(c_i) p_i$$

para toda $p \in V$.

(e) Demostrar que

$$\int_a^b p(t)dt = \sum_{i=0}^{n} p(c_i) d_i,$$

donde

$$d_i = \int_a^b p_i(t)dt.$$

Supóngase ahora que

$$c_i = a + \frac{i(b-a)}{n} \quad \text{para } i = 0, \ldots, n.$$

Para $n = 1$ la expresión anterior representa la regla trapezoidal para polinomios. Para $n = 2$, comparar el resultado con la regla de Simpson para polinomios.

11. Sean V y W espacios vectoriales dimensionalmente finitos sobre F, y sean ψ_1 y ψ_2 isomorfismos entre V y V** y W y W**, respectivamente, tal como se definieron en el Teorema 2.30. Sea T: V → W lineal, y defínase $T^{tt} = (T^t)^t$. Demostrar que el diagrama de la figura 2.4 conmuta, es decir, que $\psi_2 T = T^{tt}\psi_1$.

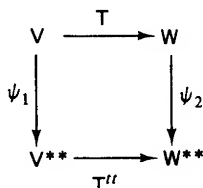


figura 2.4

12. Sea V un espacio vectorial dimensionalmente finito con la base ordenada β. Demostrar que $\psi(\beta) = \beta^{**}$, donde ψ es como se definió en el Teorema 2.30.

Para los problemas 13 a 17, V será un espacio vectorial dimensionalmente finito sobre F. Si S es un subconjunto de V, definimos al *aniquilador* S^0 de S como $S^0 = \{f \in V^*: \ f(x) = 0 \text{ para toda } x \in S\}$.

13. (a) Demostrar que S^0 es un subespacio de V*.

(b) Si W es un subespacio de V y $x \notin W$, demostrar que existe $f \in W^0$ tal que $f(x) \neq 0$.

(c) Demostrar que $S^{00} = L(\psi(S))$, donde ψ es como se definió en el Teorema 2.30.

(d) Para los subespacios W_1 y W_2, demostrar que $W_1 = W_2$ si y sólo si $W_1^0 = W_2^0$.

(e) Para los subespacios W_1 y W_2, demostrar que $(W_1 + W_2)^0 = W_1^0 \cap W_2^0$.

14. Si W es un subespacio de V, demostrar que $\dim(W) + \dim(W^0) = \dim(V)$. *Sugerencia:* Sea $\{x_1, \ldots, x_k\}$ una base ordenada de W y extiéndase a una base ordenada $\beta = \{x_1, \ldots, x_n\}$ de V. Sea $\beta^* = \{f_1, \ldots, f_n\}$. Demostrar que $\{f_{k+1}, \ldots, f_n\}$ es una base de W^0.

15. Supóngase que W es un espacio vectorial dimensionalmente finito sobre F y que $T\colon V \to W$ es lineal. Demostrar que $N(T^t) = (R(T))^0$.

16. Utilizar los Ejercicios 14 y 15 para deducir que $\text{rango}(A) = \text{rango}(A^t)$ para toda $A \in M_{m \times n}(F)$.

17. Sea $T\colon V \to V$ una transformación lineal y W un subespacio de V. Demostrar que W es T-invariante (tal como se definió en el Ejercicio 24 de la Sección 2.1) si y sólo si W^0 es T^t-invariante.

2.7* ECUACIONES DIFERENCIALES LINEALES HOMOGENEAS CON COEFICIENTES CONSTANTES

A manera de introducción a esta sección, consideremos el siguiente problema físico. Un peso de masa m se sujeta a un resorte suspendido verticalmente al que se le permite elongarse hasta que las fuerzas que obran sobre el peso están en equilibrio. Supongamos ahora que el peso permanece en reposo y superpongamos un sistema coordenado XY con el peso en el origen y el resorte localizado en la parte superior del eje Y. (Ver Fig. 2.5.)

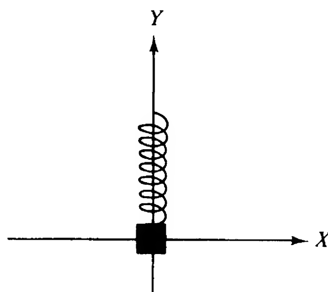


figura 2.5

Supóngase que en un cierto instante, por ejemplo $t = 0$, el peso se hace descender una distancia s a lo largo del eje Y y luego se suelta. Entonces el resorte empezará a oscilar.

Describamos el movimiento del resorte. En cualquier instante $t \geq 0$, sea $F(t)$ la fuerza que actúa sobre el peso y $y(t)$ sea la coordenada del

peso a lo largo del eje Y. Por ejemplo, $y(0) = -s$. La segunda derivada de y con respecto al tiempo, $y''(t)$, es la aceleración del peso en el instante t, y por lo tanto de acuerdo con la segunda ley de Newton

$$F(t) = my''(t). \tag{1}$$

Es razonable suponer que la fuerza que actúa sobre el peso se debe totalmente a la tensión del resorte y que esta fuerza satisface la ley de Hooke: *La fuerza que actúa sobre el peso es proporcional a su desplazamiento a partir de la posición de equilibrio, pero en dirección opuesta.* Si $k > 0$ es la constante de proporcionalidad, entonces la ley de Hooke establece que

$$F(t) = -ky(t). \tag{2}$$

Combinando las ecuaciones (1) y (2), obtenemos

$$my'' = -ky$$

o bien

$$y'' + \frac{k}{m}y = 0. \tag{3}$$

La expresión de la ecuación (3) es un ejemplo de "ecuación diferencial". Una *ecuación diferencial* en una función incógnita $y = y(t)$ es una ecuación que involucra a y, a t y a las derivadas de y. Si la ecuación diferencial es de la forma

$$a_n y^{(n)} + a_{n-1}y^{(n-1)} + \ldots + a_1 y^{(1)} + a_0 y = f, \tag{4}$$

donde $a_0, a_1, \ldots, a_n$ y f son funciones de t y $y^{(k)}$ es la k-ésima derivada de y, entonces se dice que la ecuación es *lineal*. Las funciones a_i se denominan *coeficientes* de la ecuación diferencial lineal (4). Así, la ecuación (3) es un ejemplo de ecuación diferencial lineal en donde los coeficientes son constantes y la función f es idéntica a cero. Cuando la función f de la ecuación (4) es idéntica a cero, la ecuación diferencial lineal se llama *homogénea.*

En esta sección aplicaremos el álgebra lineal que hemos estudiado para resolver ecuaciones diferenciales lineales homogéneas con coeficientes constantes. Si $a_n \neq 0$, decimos que la ecuación diferencial (4) es de orden n. En este caso podemos dividir ambos lados entre a_n para obtener una ecuación nueva pero equivalente

$$y^{(n)} + b_{n-1}y^{(n-1)} + \ldots + b_1 y^{(1)} + b_0 y = 0,$$

donde $b_i = a_i/a_n$ para $i = 0, 1, \ldots, n-1$. A causa de esta observación supondremos siempre que el coeficiente inicial a_n de la ecuación (4) es 1.

Una *solución* a la ecuación (4) es una función tal que cuando se substituye en y reduce la ecuación (4) a una identidad.

Ejemplo 37. La función $y(t) = \text{sen}\,\sqrt{k/m}t$ es una solución a la ecuación (3) puesto que

$$y''(t) + \frac{k}{m}y(t) = -\frac{k}{m}\,\text{sen}\,\sqrt{\frac{k}{m}}t + \frac{k}{m}\,\text{sen}\,\sqrt{\frac{k}{m}}t = 0$$

para toda t. Nótese, sin embargo, que al substituir $y(t) = t$ en la ecuación (3) se obtiene

$$y''(t) + \frac{k}{m}y(t) = \frac{k}{m}t,$$

lo cual no es idénticamente cero. Por tanto, $y(t) = t$ no es solución a la ecuación (3).

Al intentar resolver ecuaciones diferenciales, descubriremos que es de utilidad considerar las soluciones como funciones complejas de una variable real, aun cuando las soluciones que para nosotros tienen significado son las funciones reales de variable real. La conveniencia de este punto de vista se hará más claro posteriormente. Así, nos ocuparemos del espacio vectorial $\mathcal{F}(R, C)$ (tal como se definió en el Ejemplo 3 de la Sección 1.2). A fin de considerar a las funciones complejas de variable real como soluciones a las ecuaciones diferenciales debemos definir lo que significa diferenciar tales funciones. Dada una función compleja $x \in \mathcal{F}(R, C)$ de una variable real t, existen funciones reales únicas x_1 y x_2 de t, tales que

$$x(t) = x_1(t) + ix_2(t) \qquad \text{para } t \in R,$$

donde i es el número imaginario puro tal que $i^2 = -1$. Decimos que x_1 es la *parte real* y que x_2 es la *parte imaginaria* de x.

Definición. *Dada una función* $x \in \mathcal{F}(R, C)$ *con parte real* x_1 *y parte imaginaria* x_2, *decimos que* x *es* diferenciable *si* x_1 *y* x_2 *son diferenciables. Si* x *es diferenciable definimos la derivada de* x, x′ *como*

$$x' = x_1' + ix_2'.$$

Ejemplo 38. Si $x(t) = \cos 2t + i\,\text{sen}\,2t$, entonces

$$x'(t) = -2\,\text{sen}\,2t + i(2\cos 2t).$$

Determinamos a continuación la parte real e imaginaria de x^2. Como

$$\begin{aligned} x^2(t) &= (\cos 2t + i\,\text{sen}\,2t)^2 = (\cos^2 2t - \text{sen}^2\,2t) + i(2\,\text{sen}\,2t\cos 2t) \\ &= \cos 4t + i\,\text{sen}\,4t, \end{aligned}$$

la parte real de $x^2(t)$ es $\cos 4t$ y la parte imaginaria es $\text{sen}\,4t$.

El teorema siguiente indica que debemos limitar nuestras investigaciones a un espacio vectorial considerablemente menor que $\mathcal{F}(R, C)$. Su de-

mostración, ilustrada en el Ejemplo 39, implica un sencillo argumento de inducción, la cual omitiremos.

Teorema 2.31. *Cualquier solución a una ecuación diferencial lineal homogénea con coeficientes constantes tiene derivadas de todos los órdenes; esto es, si* x *es una solución para tal ecuación, entonces* $x^{(k)}$ *existe para todo entero positivo* k.

Ejemplo 39. Como ilustración del Teorema 2.31 considérese la ecuación

$$y^{(2)} + 4y = 0.$$

Claramente, para ser calificada como solución, una función y debe tener dos derivadas. Sin embargo, si y es una solución,

$$y^{(2)} = -4y.$$

Ahora, como $y^{(2)}$ es un múltiplo constante de una función que tiene dos derivadas, a saber la función y, $y^{(2)}$ debe tener dos derivadas y entonces $y^{(4)}$ existe. De hecho

$$y^{(4)} = -4y^{(2)}.$$

Como $y^{(4)}$ es un múltiplo constante de una función que hemos mostrado que tiene al menos dos derivadas, también tiene al menos dos derivadas, y por lo tanto $y^{(6)}$ existe. Continuando de esta manera podemos demostrar que cualquier solución tiene derivadas de todos los órdenes.

Definición. *Utilizaremos el símbolo* C^∞ *para representar al conjunto de todas las funciones en* $\mathcal{F}(R, C)$ *que tienen derivadas de todos los órdenes.*

Es un ejercicio sencillo demostrar que C^∞ es un subespacio de $\mathcal{F}(R, C)$ y, por tanto, un espacio vectorial sobre C. Como resultado del Teorema 2.31 es este espacio vectorial el que nos interesa. Para $x \in C^\infty$ la derivada x' de x también está en C^∞. Podemos utilizar la operación derivada para definir un mapeo $D\colon C^\infty \to C^\infty$ mediante

$$D(x) = x' \qquad \text{para } x \in C^\infty.$$

Es fácil demostrar que D es una transformación lineal. Más generalmente, considérese cualquier polinomio sobre C de la forma

$$p(t) = a_n t^n + a_{n-1} t^{n-1} + \dots + a_1 t + a_0.$$

Entonces

$$p(D) = a_n D^n + a_{n-1} D^{n-1} + \dots + a_1 D + a_0 I$$

es una transformación lineal. (Ver el Apéndice E.)

Definiciones. *Para cualquier polinomio* p(t) *sobre* C, p(D) *se llama* operador diferencial. *El* orden *del operador diferencial* p(D) *es el grado del polinomio* p(t).

Los operadores diferenciales son útiles porque nos proporcionan medios para reformular una ecuación diferencial dentro del contexto del álgebra lineal. Cualquier ecuación diferencial lineal homogénea con coeficientes constantes

$$y^{(n)} + a_{n-1}y^{(n-1)} + \ldots + a_1y^{(1)} + a_0y = 0$$

puede ser escrita de nuevo por medio de operadores diferenciales como

$$(\mathsf{D}^n + a_{n-1}\mathsf{D}^{n-1} + \ldots + a_1\mathsf{D} + a_0\mathsf{I})(y) = 0.$$

Definición. *Dada la ecuación diferencial anterior, el polinomio complejo*

$$\mathrm{p(t)} = \mathrm{t}^n + \mathrm{a}_{n-1}\mathrm{t}^{n-1} + \ldots + \mathrm{a}_1\mathrm{t} + \mathrm{a}_0$$

se llama polinomio auxiliar *asociado con la ecuación.*

Por ejemplo, la ecuación (3) tiene el polinomio auxiliar

$$p(t) = t^2 + \frac{k}{m}.$$

Cualquier ecuación diferencial lineal homogénea con coeficientes constantes puede reescribirse como

$$p(\mathsf{D})(y) = 0,$$

donde $p(t)$ es el polinomio auxiliar asociado con la ecuación. Claramente esta ecuación implica lo siguiente.

Teorema 2.32. *El conjunto de todas las soluciones a una ecuación diferencial lineal homogénea con coeficientes constantes coincide con el espacio nulo de* p(D), *donde* p(t) *es el polinomio auxiliar asociado con la ecuación.*

Corolario. *El conjunto de todas las soluciones a una ecuación diferencial lineal homogénea con coeficientes constantes es un subespacio de* C^∞.

En vista del corolario anterior, llamaremos al conjunto de soluciones a una ecuación diferencial lineal homogénea con coeficientes constantes, el *espacio solución* de la ecuación. Una manera práctica de describir tal espacio es encontrar una base para él. Examinaremos una cierta clase de funciones que serán de utilidad para encontrar bases para estos espacios solución.

Para un número real s, ya estamos familiarizados con el número real e^s, donde e es el número único cuyo logaritmo natural es 1 (esto es,

$\ln(e) = 1$). Por ejemplo, conocemos algunas propiedades de la exponenciación:

$$e^{s+t} = e^s e^t \quad \text{y} \quad e^{-t} = \frac{1}{e^t}$$

para cualquier par de números reales s y t. Extenderemos ahora la definición de las potencias de e para incluir a los números complejos de tal modo que estas propiedades continúen siendo válidas.

Definición. *Sea* $c = a + ib$ *cualquier número complejo con parte real* a *y parte imaginaria* b. *Defínase*

$$e^c = e^a(\cos b + i \operatorname{sen} b).$$

Por ejemplo, para $c = 2 + i(\pi/3)$,

$$e^c = e^2\left(\cos\frac{\pi}{3} + i \operatorname{sen}\frac{\pi}{3}\right) = e^2\left(\frac{1}{2} + i\frac{\sqrt{3}}{2}\right).$$

Claramente, si c es real ($b = 0$), obtenemos el resultado usual $e^c = e^a$. Puede demostrarse con el uso de identidades trigonométricas que

$$e^{c+d} = e^c e^d \quad \text{y} \quad e^{-c} = \frac{1}{e^c}$$

para cualquier par de complejos c y d.

Definición. *Sea* c *cualquier número complejo. La función* $f\colon R \to C$ *definida por* $f(t) = e^{ct}$ *para toda* t *en* R *se llama* función exponencial.

La derivada de una función exponencial, como se describe en el teorema siguiente, es como esperaríamos. La demostración implica un cálculo directo, aunque tedioso, que dejaremos como ejercicio.

Teorema 2.33. *Para cualquier función exponencial* $f(t) = e^{ct}$, $f'(t) = ce^{ct}$.

Utilizaremos funciones exponenciales para describir todas las soluciones de una ecuación diferencial lineal homogénea de orden 1. Recuérdese que el *orden* de dicha ecuación es igual al grado de su polinomio auxiliar, de manera que una ecuación de orden 1 es de la forma

$$y' + a_0 y = 0 \tag{5}$$

Teorema 2.34. *El espacio solución para la ecuación (5) es de dimensión 1 y tiene a* $\{e^{-a_0 t}\}$ *como base.*

DEMOSTRACIÓN. Claramente, la ecuación (5) tiene a $e^{-a_0 t}$ como solución. Supóngase que $x(t)$ es una solución cualquiera a la ecuación (5). Entonces

$$x'(t) = -a_0 x(t) \qquad \text{para toda } t \in R.$$

Defínase

$$z(t) = e^{a_0 t}x(t).$$

Al diferenciar z obtenemos

$$z'(t) = (e^{a_0 t})'x(t) + e^{a_0 t}x'(t) = a_0 e^{a_0 t}x(t) - a_0 e^{a_0 t}x(t) = 0.$$

Nótese que la conocida regla para la diferenciación de productos se conserva para funciones complejas de variable real. Una justificación implica un cálculo directo aunque bastante largo.

Como z' es idénticamente cero, z es una función constante. De nuevo, este hecho, muy conocido para funciones reales de variable real, también es cierto para funciones complejas; la demostración, semejante a la que se hace para el caso real, implica considerar por separado las partes real e imaginaria de z. Entonces, existe un número complejo c tal que

$$z(t) = e^{a_0 t}x(t) = c \qquad \text{para toda } t \in R.$$

Así,

$$x(t) = ce^{-a_0 t}.$$

De donde concluimos que cualquier miembro del espacio solución de la ecuación (5) es una combinación lineal de $e^{-a_0 t}$. ■

Otra manera de formular el Teorema 2.34 es la siguiente.

Corolario. *Para cualquier número complejo* c, *el espacio nulo del operador diferencial* $\mathsf{D} - c\mathsf{I}$ *tiene como base a* $\{e^{ct}\}$.

Nos ocuparemos ahora de ecuaciones diferenciales de orden superior a uno. Dada una ecuación diferencial lineal homogénea de orden n con coeficientes constantes

$$y^{(n)} + a_{n-1}y^{(n-1)} + \ldots + a_1 y^{(1)} + a_0 y = 0,$$

su polinomio auxiliar

$$p(t) = t^n + a_{n-1}t^{n-1} + \ldots + a_1 t + a_0$$

se descompone en un producto de factores de grado 1:

$$p(t) = (t - c_1)(t - c_2) \ldots (t - c_n),$$

donde $c_1, c_2, \ldots, c_n$ son números complejos (no necesariamente distintos). (Esto se deduce del teorema fundamental del álgebra dado en el Apéndice D.) Entonces

$$p(\mathsf{D}) = (D - c_1\mathsf{I})(\mathsf{D} - c_2\mathsf{I}) \ldots (\mathsf{D} - c_n\mathsf{I}).$$

Pero los operadores $\mathsf{D} - c_i\mathsf{I}$ conmutan y así, por el Ejercicio 9, tenemos que

$$\mathsf{N}(\mathsf{D} - c_i\mathsf{I}) \subset \mathsf{N}(p(\mathsf{D})) \qquad \text{para toda } i.$$

Como $N(p(D))$ coincide con el espacio solución de la ecuación diferencial dada, podemos concluir el siguiente resultado por el corolario del Teorema 2.34.

Teorema 2.35. *Sea* $p(t)$ *el polinomio auxiliar para una ecuación diferencial lineal homogénea con coeficientes constantes. Para cualquier número complejo* c, *si* c *es un cero de* $p(t)$, *entonces* e^{ct} *es una solución a la ecuación diferencial.*

Ejemplo 40. Dada la ecuación diferencial

$$y'' - 3y' + 2y = 0,$$

su polinomio auxiliar $p(t) = t^2 - 3t + 2$ se puede factorizar como

$$p(t) = (t - 1)(t - 2).$$

Por tanto, por el Teorema 2.35, e^t y e^{2t} son soluciones de la ecuación anterior porque $c = 1$ y $c = 2$ son ceros de $p(t)$. Como el espacio solución de la ecuación anterior es un subespacio de C^∞, $L(\{e^t, e^{2t}\})$ se encuentra en el espacio solución. Es sencillo demostrar que $\{e^t, e^{2t}\}$ es linealmente independiente y si pudiéramos demostrar que el espacio solución es bidimensional, podríamos concluir que $\{e^t, e^{2t}\}$ es una base para el espacio solución. Este resultado se deduce del teorema siguiente.

Teorema 2.36. *Para cualquier operador diferencial* $p(D)$ *de orden* n, *el espacio nulo de* $p(D)$ *es un subespacio* n*-dimensional de* C^∞.

Como preliminares de la demostración del Teorema 2.36 estableceremos dos lemas.

Lema 1. *El operador diferencial* $D - cI$: $C^\infty \to C^\infty$ *es sobreyectivo para cualquier número complejo* C.

DEMOSTRACIÓN. Sea $x \in C^\infty$. Deseamos encontrar un $y \in C^\infty$ tal que $(D - cI)y = x$. Defínase una función w mediante $w(t) = x(t)e^{-ct}$ para $t \in R$.

Claramente, $w \in C^\infty$ pues x y e^{-ct} están en C^∞. Sean respectivamente w_1 y w_2 las partes real e imaginaria de w. Como $w \in C^\infty$, w_1 y w_2 son diferenciables y por tanto continuas, por lo que tienen antiderivadas, digamos que son W_1 y W_2, tales que $W_1' = w_1$ y $W_2' = w_2$. Defínase W: $R \to C$ mediante

$$W(t) = W_1(t) + iW_2(t) \qquad \text{para } t \in R.$$

Entonces $W \in C^\infty$ y las partes real e imaginaria de W son W_1 y W_2, respectivamente. También $W' = w$. Finalmente, defínase y: $R \to C$ mediante $y(t) = W(t)e^{ct}$ para $t \in R$.

Claramente, $y \in C^\infty$ y como

$$\begin{aligned}(\mathsf{D} - c\mathsf{I})y(t) &= y'(t) - cy(t)\\ &= W'(t)e^{ct} + W(t)ce^{ct} - cW(t)e^{ct}\\ &= w(t)e^{ct}\\ &= x(t)e^{-ct}e^{ct}\\ &= x(t),\\ (\mathsf{D} - c\mathsf{I})y &= x. \quad \blacksquare\end{aligned}$$

Lema 2. *Sea* V *un espacio vectorial y supóngase que* T *y* U *son operadores lineales en* V *tales que*

(a) U *es sobreyectiva.*

(b) *Los espacios nulos de* T *y* U *son dimensionalmente finitos. Entonces el espacio nulo de* TU *es dimensionalmente finito y*

$$dim(\mathsf{N}(\mathsf{TU})) = dim(\mathsf{N}(\mathsf{T})) + dim(\mathsf{N}(\mathsf{U})).$$

DEMOSTRACIÓN. Sean $p = \dim(\mathsf{N}(\mathsf{T}))$, $q = \dim(\mathsf{N}(\mathsf{U}))$ y $\{u_1, u_2, \ldots, u_p\}$ y $\{v_1, v_2, \ldots, v_q\}$ sean bases para N(T) y N(U), respectivamente. Como U es sobreyectiva podemos seleccionar para cada $i (i = 1, \ldots, p)$ un elemento $w_i \in \mathsf{V}$ tal que $\mathsf{U}(w_i) = u_i$. Entonces obtenemos un conjunto de p elementos $\{w_1, w_2, \ldots, w_p\}$. Nótese que para cualquier i y j, $w_i \neq v_j$ puesto que de otra manera $u_i = \mathsf{U}(w_i) = \mathsf{U}(v_j) = 0$ lo que es una contradicción. Por lo tanto, el conjunto

$$\beta = \{w_1, w_2, \ldots, w_p, v_1, \ldots, v_q\}$$

contiene $p + q$ elementos diferentes. Para demostrar el lema es suficiente demostrar que β es una base para N(TU).

Demostraremos primero que β genera a N(TU). Como para toda w_i y v_j en β

$$\mathsf{TU}(w_i) = \mathsf{T}(u_i) = 0 \quad \text{y} \quad \mathsf{TU}(v_j) = \mathsf{T}(0) = 0,$$

$$\beta \subseteq \mathsf{N}(\mathsf{TU}).$$

Ahora supóngase que $v \in \mathsf{N}(\mathsf{TU})$. Entonces

$$0 = \mathsf{TU}(v) = \mathsf{T}(\mathsf{U}(v)).$$

Así, tenemos que $\mathsf{U}(v) \in \mathsf{N}(\mathsf{T})$ y existirán escalares $a_1, a_2, \ldots, a_p$ tales que

$$\begin{aligned}\mathsf{U}(v) &= a_1u_1 + a_2u_2 + \ldots + a_pu_p\\ &= \mathsf{U}(a_1w_1 + a_2w_2 + \ldots + a_pw_p).\end{aligned}$$

Por lo tanto

$$\mathsf{U}(v - (a_1w_1 + a_2w_2 + \ldots + a_pw_p)) = 0.$$

Concluimos que $v - (a_1w_1 + \ldots + a_pw_p)$ está en N(U). De aquí se sigue que existen escalares $b_1, b_2, \ldots, b_q$ tales que

$$v - (a_1w_1 + a_2w_2 + \ldots + a_pw_p) = b_1v_1 + b_2v_2 + \ldots + b_qv_q$$

o

$$v = a_1w_1 + a_2w_2 + \ldots + a_pw_p + b_1v_1 + b_2v_2 + \ldots + b_qv_q.$$

Por lo tanto β genera a N(TU).

Demostraremos ahora que β es linealmente independiente. Sean $a_1, a_2, \ldots, a_p, b_1, b_2, \ldots, b_q$ escalares arbitrarios tales que

$$a_1w_1 + a_2w_2 + \ldots + a_pw_p + b_1v_1 + b_2v_2 + \ldots + b_qv_q = 0. \qquad (6)$$

Aplicando U en ambos lados de la ecuación (6), obtenemos

$$a_1u_1 + a_2u_2 + \ldots + a_pu_p = 0.$$

Como $\{u_1, u_2, \ldots, u_p\}$ es linealmente independiente, todas las a_i son cero. Así, la ecuación (6) se reduce a

$$b_1v_1 + b_2v_2 + \ldots + b_qv_q = 0.$$

Ahora, la independencia lineal de $\{v_1, v_2, \ldots, v_q\}$ implica que las b_i son todas cero. Concluimos entonces que β es una base para N(TU). Por lo tanto, N(TU) es dimensionalmente finito y $\dim(\mathsf{N}(\mathsf{TU})) = p + q = \dim(\mathsf{N}(\mathsf{T})) + \dim(\mathsf{N}(\mathsf{U}))$. ■

DEMOSTRACIÓN DEL TEOREMA 2.36. La demostración requerirá de un argumento de inducción matemática sobre el orden del operador diferencial $p(\mathsf{D})$. El caso de primer orden coincide con el Teorema 2.34. Entonces, para algún entero $n > 1$ supóngase que el Teorema 2.36 se satisface para cualquier operador diferencial de orden menor que n y supóngase que se tiene un operador diferencial $p(\mathsf{D})$ de orden n. El polinomio $p(t)$ puede descomponerse en un producto de dos polinomios

$$p(t) = q(t)(t - c)$$

para algún polinomio $q(t)$ de grado $n - 1$ y para algún número complejo c. Entonces, el operador diferencial dado puede reescribirse como

$$p(\mathsf{D}) = q(\mathsf{D})(\mathsf{D} - c\mathsf{I}).$$

De acuerdo con el Lema 1, $\mathsf{D} - c\mathsf{I}$ es sobreyectivo; por el corolario del Teorema 2.34 $\dim(\mathsf{N}(\mathsf{D} - c\mathsf{I})) = 1$; y por la hipótesis de inducción $\dim(\mathsf{N}(q(\mathsf{D})) = n - 1$. Luego, al aplicar el Lema 2 concluimos que

$$\begin{aligned}\dim(\mathsf{N}(p(\mathsf{D}))) &= \dim(\mathsf{N}(q(\mathsf{D}))) + \dim(\mathsf{N}(\mathsf{D} - c\mathsf{I})) \\ &= (n - 1) + 1 = n.\end{aligned}$$ ■

Corolario. *Para cualquier ecuación diferencial lineal homogénea con coeficientes constantes, el espacio solución es un subespacio* n*-dimensional de* C^∞.

El corolario del Teorema 2.36 reduce el problema de encontrar todas las soluciones de una ecuación diferencial lineal homogénea de coeficientes constantes de orden n al problema de encontrar el conjunto de n soluciones linealmente independientes de la ecuación. Por los resultados del Capítulo 1, cualquier conjunto tal debe ser una base para el espacio solución. El teorema siguiente nos permite encontrar rápidamente una base para muchas de estas ecuaciones. En los ejercicios se dan algunas sugerencias para su demostración.

Teorema 2.37. *Dados* n *números complejos distintos* $c_1, c_2, \ldots, c_n$, *el conjunto de funciones exponenciales* $\{e^{c_1 t}, e^{c_2 t}, \ldots, e^{c_n t}\}$ *es linealmente independiente.*

Corolario. *Para cualquier ecuación diferencial lineal homogénea con coeficientes constantes de orden* n, *si su polinomio auxiliar* p(t) *tiene* n *ceros distintos* $c_1, c_2, \ldots, c_n$, *entonces el conjunto* $\{e^{c_1 t}, e^{c_2 t}, \ldots, e^{c_n t}\}$ *es una base para el espacio solución de la ecuación.*

DEMOSTRACIÓN. Ejercicio.

Ejemplo 41. Encontremos todas las soluciones a la ecuación diferencial

$$y'' + 5y' + 4y = 0.$$

Como el polinomio auxiliar $p(t)$ se puede factorizar como $(t + 4)(t + 1)$, $p(t)$ tiene dos ceros distintos: -1 y -4. Entonces, $\{e^{-t}, e^{-4t}\}$ es una base para el espacio solución, y entonces cualquier solución para la ecuación dada es de la forma

$$y(t) = b_1 e^{-t} + b_2 e^{-4t} \qquad \text{para algunas constantes } b_1 \text{ y } b_2.$$

Ejemplo 42. Encontremos todas las soluciones para la ecuación diferencial

$$y'' + 9y = 0.$$

El polinomio auxiliar $p(t) = t^2 + 9$ puede factorizarse como $p(t) = (t - 3i)(t + 3i)$ y por lo tanto tiene ceros distintos $c_1 = 3i$, $c_2 = -3i$. Así, $\{e^{3it}, e^{-3it}\}$ es una base para el espacio solución. Una base de mucho mayor utilidad podría obtenerse aplicando el Ejercicio 7. Como

$$\cos 3t = \frac{1}{2}(e^{3it} + e^{-3it}) \quad \text{y} \quad \operatorname{sen} 3t = \frac{1}{2i}(e^{3it} - e^{-3it}),$$

se deduce que {cos $3t$, sen $3t$} también es una base. Esta base tiene la ventaja sobre la original de que consiste de las conocidas funciones seno y coseno y no hace referencia al número imaginario i.

Ahora considérese la ecuación diferencial

$$y'' + 2y' + y = 0,$$

para la cual el polinomio auxiliar es $p(t) = (t + 1)^2$. Por el Teorema 2.35, e^{-t} es una solución a la ecuación anterior. Por el corolario del Teorema 2.36 su espacio solución es bidimensional. Con el objeto de encontrar una base para el espacio solución, necesitamos encontrar una solución que sea linealmente independiente de e^{-t}. El lector podrá verificar que te^{-t} cumple con esta condición y entonces $\{e^{-t}, te^{-t}\}$ es una base para el espacio solución. Este resultado puede generalizarse de la siguiente manera.

Teorema 2.38. *Sea* $p(t) = (t - c)^n$, *donde* c *es un número complejo y* n *es un entero positivo, el polinomio auxiliar de una ecuación diferencial lineal homogénea con coeficientes constantes. El conjunto*

$$\beta = \{e^{ct}, te^{ct}, \ldots, t^{n-1}e^{ct}\}$$

es una base para el espacio solución.

DEMOSTRACIÓN. Como el espacio solución es n-dimensional necesitamos únicamente demostrar que β es linealmente independiente y que está en el espacio solución. Primero obsérvese que para cualquier entero positivo k

$$\begin{aligned}(\mathsf{D} - c\mathsf{I})(t^k e^{ct}) &= kt^{k-1}e^{ct} + ct^k e^{ct} - ct^k e^{ct}\\ &= kt^{k-1}e^{ct}.\end{aligned}$$

Por lo tanto, para $k < n$,

$$(\mathsf{D} - c\mathsf{I})^n(t^k e^{ct}) = 0.$$

Y se tiene que β es un subconjunto del espacio solución.

Ahora demostraremos que β es linealmente independiente. Considérese cualquier combinación de β tal que

$$b_1 t^{n-1}e^{ct} + b_2 t^{n-2}e^{ct} + \ldots + b_{n-1}te^{ct} + b_n e^{ct} = 0 \tag{8}$$

para algunos escalares $b_1, \ldots, b_n$. Dividiendo la ecuación (8) por e^{ct}, obtenemos

$$b_1 t^{n-1} + b_2 t^{n-2} + \ldots + b_{n-1}t + b_n = 0. \tag{9}$$

Por lo tanto, el lado izquierdo de la ecuación (9) debe ser la función polinomial cero, de donde concluimos que los coeficientes $b_1, b_2, \ldots, b_n$ son todos cero. Luego, β es linealmente independiente y por lo tanto es una base para el espacio solución. ■

Ejemplo 43. Dada la ecuación diferencial

$$y^{(4)} - 4y^{(3)} + 6y^{(2)} - 4y^{(1)} + y = 0,$$

deseamos encontrar una base para el espacio de soluciones. Como su polinomio auxiliar es

$$p(t) = t^4 - 4t^3 + 6t^2 - 4t + 1 = (t-1)^4,$$

podemos concluir por el Teorema 2.38 que $\{e^t, te^t, t^2e^t, t^3e^t\}$ es una base para el espacio de soluciones, de manera que cualquier solución a la ecuación dada es de la forma

$$y(t) = b_1e^t + b_2te^t + b_3t^2e^t + b_4t^3e^t$$

para algunos escalares b_1, b_2, b_3 y b_4.

La situación más general (cuya demostración dejamos como ejercicio) puede enunciarse de la siguiente manera.

Teorema 2.39. *Para una ecuación diferencial lineal con coeficientes constantes cuyo polinomio auxiliar es*

$$p(t) = (t-c_1)^{n_1}(t-c_2)^{n_2}\dots(t-c_k)^{n_k},$$

donde $n_1, n_2, \dots, n_k$ *son enteros positivos y* $c_1, c_2, \dots, c_k$ *son números complejos distintos, el siguiente conjunto es una base para el espacio solución de la ecuación:*

$$\{e^{c_1t}, te^{c_1t}, \dots, t^{n_1-1}e^{c_1t}, \dots, e^{c_kt}, te^{c_kt}, \dots, t^{n_k-1}e^{c_kt}\}.$$

Ejemplo 44. Considérese la ecuación diferencial

$$y^{(3)} - 4y^{(2)} + 5y^{(1)} - 2y = 0.$$

Encontraremos una base para su espacio solución. Como el polinomio auxiliar $p(t)$ puede factorizarse como

$$p(t) = t^3 - 4t^2 + 5t - 2 = (t-1)^2(t-2),$$

concluimos que el espacio solución de la ecuación diferencial anterior tiene como base

$$\{e^t, te^t, e^{2t}\}.$$

Por ello cualquier solución a la ecuación dada es de la forma

$$y(t) = b_1e^t + b_2te^t + b_3e^{2t}$$

para algunos escalares b_1, b_2 y b_3.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

   (a) El conjunto de soluciones de una ecuación diferencial lineal homogénea de orden n con coeficientes constantes es un subespacio n-dimensional de C^{∞}.
   (b) El espacio de soluciones de una ecuación diferencial lineal homogénea es el espacio nulo de un operador diferencial.
   (c) El polinomio auxiliar de una ecuación diferencial lineal homogénea con coeficientes constantes es una solución a la ecuación diferencial.
   (d) Cualquier solución a una ecuación diferencial lineal homogénea con coeficientes constantes es de la forma ae^{ct} o $at^k e^{ct}$ donde a y c son números complejos y k es un entero positivo.
   (e) Cualquier combinación lineal de soluciones a una ecuación diferencial lineal homogénea con coeficientes constantes es también una solución a la ecuación dada.
   (f) Para cualquier ecuación diferencial lineal homogénea con coeficientes constantes que tenga un polinomio auxiliar $p(t)$, si $c_1, c_2, \ldots, c_k$ son las distintas raíces de $p(t)$, entonces $\{e^{c_1t}, e^{c_2t}, \ldots, e^{c_kt}\}$ es una base para el espacio de soluciones de la ecuación diferencial dada.
   (g) Dado cualquier polinomio $p(t) \in P(C)$, existe una ecuación diferencial lineal homogénea con coeficientes constantes cuyo polinomio auxiliar es $p(t)$.

2. Para cada uno de los incisos siguientes, determinar si el enunciado es verdadero o falso. Justificar la respuesta con una demostración o en su caso con un contraejemplo.

   (a) Cualquier subespacio dimensionalmente finito de C^{∞} es el espacio solución de una ecuación diferencial lineal homogénea con coeficientes constantes.
   (b) Existe una ecuación diferencial lineal homogénea con coeficientes constantes cuyo espacio solución tiene como base a $\{t, t^2\}$.
   (c) Para cualquier ecuación diferencial lineal homogénea con coeficientes constantes, si x es una solución, también lo será su derivada x'.

   Dados dos polinomios $p(t)$ y $g(t)$ en $P(C)$, si $x \in N(p(D))$ y $y \in N(q(D))$ entonces

   (d) $x + y \in N(p(D)q(D))$.
   (e) $xy \in N(p(D)q(D))$.

3. Encontrar bases para los espacios de soluciones de las siguientes ecuaciones diferenciales.

(a) $y'' + 2y' + y = 0$
(b) $y''' = y'$
(c) $y^{(4)} - 2y^{(2)} + y = 0$
(d) $y'' + 2y' + y = 0$
(e) $y^{(3)} - y^{(2)} + 3y^{(1)} + 5y = 0$

4. Encontrar bases para los siguientes subespacios de C^∞.

(a) $N(D^2 - D - I)$
(b) $N(D^3 - 3D^2 + 3D - I)$
(c) $N(D^3 + 6D^2 + 8D)$

5. Demostrar que C^∞ es un subespacio de $\mathcal{F}(R, C)$.

6. (a) Probar que D: $C^\infty \to C^\infty$ es una transformación lineal.
(b) Probar que cualquier operador diferencial es una transformación lineal en C^∞.

7. Demostrar que si $\{x, y\}$ es una base para un espacio vectorial sobre C, entonces también lo es

$$\left\{\frac{1}{2}(x + y), \frac{1}{2i}(x - y)\right\}.$$

8. Dada una ecuación diferencial lineal homogénea con coeficientes constantes de segundo orden, supóngase que el polinomio auxiliar tiene raíces complejas conjugadas $a + ib$ y $a - ib$, donde $a, b \in R$. Demostrar que $\{e^{at} \cos bt, e^{at} \operatorname{sen} bt\}$ es una base para el espacio solución.

9. Dada una colección de transformaciones lineales conmutativas por parejas $\{U_1, U_2, \ldots, U_n\}$ de un espacio vectorial V (es decir, transformaciones tales que $U_iU_j = U_jU_i$ para toda i, j), probar que para cualquier $i = 1, 2, \ldots, n$.

$$N(U_i) \subseteq N(U_1U_2 \ldots U_n).$$

10. Demostrar el Teorema 2.37 y su corolario. *Sugerencia:* Suponer que

$$b_1e^{c_1t} + b_2e^{c_2t} + \ldots + b_ne^{c_nt} = 0 \quad \text{(donde los } c_i \text{ son distintos).}$$

Para demostrar que los b_i son cero, aplicar inducción matemática sobre n. Verificar el teorema para $n = 1$. Suponiendo que el teorema es cierto para cualquier $n - 1$ funciones, aplicar el operador $D - c_nI$ a ambos lados de la ecuación anterior para establecer el teorema para n diferentes funciones exponenciales.

11. Demostrar el Teorema 2.39. *Sugerencia:* Primero verifíquese que la base supuesta se encuentra en el espacio solución. Luego verifíquese que este conjunto es linealmente independiente por inducción matemática sobre k.

El caso $k = 1$ es el Teorema 2.38. Suponiendo que el teorema se cumple para $k - 1$ c_i diferentes, aplíquese el operador $(\mathsf{D} - c_k\mathsf{I})^{n_k}$ a cualquier combinación lineal de la base supuesta que sea igual a cero.

12. Sea V el espacio de soluciones de una ecuación diferencial lineal homogénea de orden n con coeficientes constantes, cuyo polinomio auxiliar es $p(t)$. Demostrar que si $p(t) = g(t)h(t)$, donde $g(t)$ y $h(t)$ son polinomios de grado positivo, entonces

$$\mathsf{N}(h(\mathsf{D})) = \mathsf{R}(g(\mathsf{D}_\mathsf{V})) = g(\mathsf{D})(\mathsf{V}),$$

donde D_V: $\mathsf{V} \to \mathsf{V}$ está definida mediante $\mathsf{D}_\mathsf{V}(x) = x'$ para $x \in \mathsf{V}$. *Sugerencia:* Demostrar primero que $g(\mathsf{D})(\mathsf{V}) \subseteq \mathsf{N}(h(\mathsf{D}))$. Entonces, probar que los dos espacios tienen la misma dimensión finita.

13. Una ecuación diferencial

$$y^{(n)} + a_{n-1}y^{(n-1)} + \ldots + a_1y^{(1)} + a_0y = x$$

se denomina ecuación diferencial lineal *no homogénea* con coeficientes constantes si los coeficientes a_i son constantes y el lado derecho de la ecuación, x, es una función que no es idénticamente nula.

(a) Demostrar que para cualquier $x \in \mathsf{C}^\infty$ existe una $y \in \mathsf{C}^\infty$ tal que y es una solución para la ecuación anterior. *Sugerencia:* Utilizar el Lema 1 del Teorema 2.36 para demostrar que si

$$p(t) = t^n + a_{n-1}t^{n-1} + \ldots + a_1t + a_0,$$

entonces $p(\mathsf{D})$: $\mathsf{C}^\infty \to \mathsf{C}^\infty$ es sobreyectiva.

(b) Sea V el espacio de soluciones para la ecuación lineal homogénea

$$y^{(n)} + a_{n-1}y^{(n-1)} + \ldots + a_1y^{(1)} + a_0y = 0.$$

Demostrar que si z es cualquier solución a la ecuación diferencial lineal no homogénea anterior, entonces el conjunto de todas las soluciones a la ecuación diferencial lineal no homogénea es

$$\{z + y:\ y \in \mathsf{V}\}.$$

14. Dada cualquier ecuación diferencial lineal no homogénea de orden n con coeficientes constantes, demostrar que para cualquier solución x y cualquier $t_0 \in R$ si $x(t_0) = x'(t_0) = \ldots = x^{(n-1)}(t_0) = 0$, entonces $x = 0$ (la función cero). *Sugerencia:* Emplear inducción matemática sobre n. Primero demostrar la conclusión para el caso $n = 1$. Luego, supóngase que es cierto para ecuaciones de orden $n - 1$ y considérese una ecuación de orden n con polinomio auxiliar $p(t)$. Descomponer $p(t)$ como $p(t) = q(t)(t - c)$ para algún número complejo c y un polinomio $q(t)$ de grado $n - 1$. Sea $z = q(\mathsf{D})x$. Demostrar que $z(t_0) = 0$ y que z es una solución para la ecuación $y' - cy = 0$. Concluir que $z = 0$. Ahora aplíquese la hipótesis de inducción.

15. Sea V el espacio solución de una ecuación diferencial lineal homogénea de orden n con coeficientes constantes. Fijar $t_0 \in R$ y definir un mapeo Φ: $V \to C^n$ por

$$\Phi(x) = \begin{pmatrix} x(t_0) \\ x'(t_0) \\ \cdot \\ \cdot \\ \cdot \\ x^{(n+1)}(t_0) \end{pmatrix} \quad \text{para cada } x \in V.$$

(a) Demostrar que Φ es lineal y que su espacio nulo es trivial. Deducir que Φ es un isomorfismo. *Sugerencia:* Usar el Ejercicio 14.

(b) Demostrar lo siguiente: Para cualquier ecuación diferencial lineal homogénea de orden n con coeficientes constantes, cualquier $t_0 \in R$ y cualesquiera números complejos $c_0, c_1, \ldots, c_{n-1}$ (no necesariamente distintos), existe exactamente una solución, x, a la ecuación diferencial dada tal que $x^{(k)}(t_0) = c_k$ para $k = 0, 1, \ldots, n-1$.

16. *Movimiento pendular.* Es bien sabido que el movimiento de un péndulo se puede representar por la ecuación diferencial

$$\theta'' + \frac{g}{l}\theta = 0,$$

donde $\theta(t)$ es el ángulo en radianes que el péndulo forma con una línea vertical en el tiempo t (ver Fig. 2.6) interpretado de tal modo que θ sea positivo si el péndulo está a la derecha y negativo si el péndulo está a la izquierda de la línea vertical, desde el punto de vista del lector. En este

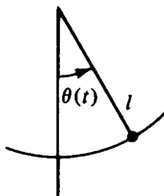


figura 2.6

caso l es la longitud del péndulo y g es la magnitud de la aceleración debida a la gravedad. La variable t y las constantes l y g deben estar expresadas en unidades compatibles, por ejemplo, t en segundos, l en metros y g en metros por segundo por segundo.

(a) Expresar una solución arbitraria a esta ecuación como una combinación lineal de dos funciones reales fijas.

(b) Encontrar la solución única de la ecuación que satisface las condiciones

$$\theta(0) = \theta_0 > 0 \quad \text{y} \quad \theta'(0) = 0.$$

(El significado de lo anterior es que en un tiempo $t = 0$ el péndulo está desplazado de la posición vertical θ_0 radianes y tiene una velocidad cero.)

(c) Demostrar que al péndulo le toma $2\pi\sqrt{\frac{l}{g}}$ unidades de tiempo para completar un recorrido completo de ida y vuelta. (Este lapso se denomina *período* del péndulo.)

17. *Movimiento Periódico de un Resorte con Amortiguamiento.* Al principio de esta sección discutimos el movimiento de un resorte que oscila bajo la suposición de que la única fuerza que actuaba sobre el resorte era la fuerza debida a la tensión del mismo. Encontramos en este caso que la ecuación (3) describía el movimiento del resorte.

(a) Encontrar la forma general de todas las soluciones de la ecuación (3).

Si analizamos el comportamiento de la solución general del inciso (a), vemos que la solución es una función periódica. Por tanto la ecuación (3) indica que el resorte nunca cesará de oscilar. Sin embargo, sabemos por experiencia, que la amplitud de la oscilación decrece hasta que finalmente el movimiento cesa. La razón por la cual la solución del inciso (a) no explica este comportamiento es que hemos ignorado el efecto de la fricción sobre el peso en movimiento. A bajas velocidades, tales como la que se está considerando, la resistencia del aire proporciona un ejemplo de amortiguamiento viscoso —la resistencia es proporcional a la velocidad del peso en movimiento pero en dirección opuesta. Para hacer una corrección a causa de la resistencia del aire, debemos añadir a la ecuación (2) el término $-ry'$. La constante $r > 0$ depende del medio en el cual ocurre el movimiento (en este caso el aire), y el término $-ry'$ tiene un signo negativo debido a que la resistencia tiene siempre sentido opuesto al del movimiento. Luego entonces, la ecuación diferencial del movimiento es $my'' = -ry' - ky$; es decir,

$$my'' + ry' + ky = 0.$$

(b) Encontrar la solución general de esta ecuación.
(c) Encontrar la solución única del inciso (b) que satisface las condiciones iniciales $y(0) = 0$ y $y'(0) = v_0$.
(d) Para $y(t)$ del inciso (c), demostrar que la amplitud de la oscilación decrece hasta llegar a cero; o sea, demostrar que

$$\lim_{t\to\infty} y(t) = 0.$$

18. Al principio de esta sección se enunció que es útil considerar las soluciones de las ecuaciones diferenciales como funciones complejas de variable

real, aun cuando las soluciones que para nosotros tienen significado en el sentido físico son funciones reales. Justificar este punto de vista.

19. El siguiente conjunto de ejercicios no implican el uso de álgebra lineal; los enunciamos sólo para completar un poco más el tema.

(a) Demostrar el Teorema 2.31. *Sugerencia:* Utilizar inducción matemática sobre el número de derivadas que tiene una solución.

(b) Demostrar (i) $e^{c+d} = e^c e^d$.

(ii) $e^{-c} = \dfrac{1}{e^c}$ para $c,\ d \in C$.

(c) Demostrar el Teorema 2.33.

(d) Verificar la regla del producto de la diferenciación para funciones complejas de variable real: Para cualquier par de funciones diferenciables x y y en $\mathcal{F}(R, C)$ el producto xy es diferenciable y

$$(xy)' = x'y + xy'.$$

Sugerencia: Encontrar las partes reales e imaginarias de xy en términos de las x y las de y, luego proceder con la diferenciación.

(e) Demostrar que si $x \in \mathcal{F}(R, C)$ y $x' = 0$, entonces x es una función constante.

INDICE DE DEFINICIONES PARA EL CAPITULO 2

Capítulo 3

Operaciones elementales en matrices y sistemas de ecuaciones lineales

Este capítulo está dedicado al logro de dos objetivos relacionados entre sí:

1. El estudio de algunas operaciones "que conservan el rango" en las matrices.
2. La aplicación de estas operaciones y de la teoría de las transformaciones lineales a la solución de sistemas de ecuaciones lineales.

Como una consecuencia del primer objetivo obtendremos un método sencillo para calcular el rango de una transformación lineal entre espacios vectoriales dimensionalmente finitos, aplicando las operaciones que conservan el rango a una matriz que representa dicha transformación.

La solución de sistemas de ecuaciones lineales es probablemente la aplicación más importante del álgebra lineal. El conocido método de eliminación para resolver sistemas de ecuaciones lineales, que fue discutido en la Sección 1.4, implica la eliminación de variables de manera que se pueda obtener un sistema más sencillo. Esta técnica por medio de la cual se eliminan las variables utiliza tres tipos de operaciones:

1. Intercambio de dos ecuaciones cualesquiera del sistema.
2. Multiplicación de cualquier ecuación del sistema por una constante no nula.
3. Suma de un múltiplo de una ecuación a otra ecuación.

Veremos en la Sección 3.3 que un sistema de ecuaciones lineales puede ser expresado como una ecuación matricial sencilla. En esta representación del sistema las tres operaciones mencionadas anteriormente son las "operaciones elementales con los renglones" en las matrices. Estas operaciones proporcionarán un método de cálculo conveniente para determinar todas las soluciones de un sistema de ecuaciones lineales.

3.1 OPERACIONES ELEMENTALES EN MATRICES Y MATRICES ELEMENTALES

En esta sección definiremos las operaciones elementales para matrices que serán empleadas a lo largo de este capítulo. En las secciones subsecuentes estas operaciones serán utilizadas para obtener métodos sencillos de cálculo, para determinar el rango de una transformación lineal y las soluciones de un sistema de ecuaciones lineales. Existen dos tipos de operaciones matriciales elementales —operaciones sobre los renglones y operaciones sobre las columnas. Como veremos a continuación, las operaciones con los renglones son de mayor utilidad. Estas operaciones surgen de las tres operaciones que se pueden emplear para eliminar variables en un sistema de ecuaciones lineales.

Sea A una matriz de $m \times n$ en un campo F. Recuérdese que A puede considerarse como un arreglo de m renglones,

$$A = \begin{pmatrix} A_1 \\ A_2 \\ \cdot \\ \cdot \\ \cdot \\ A_m \end{pmatrix}$$

o como un arreglo de n columnas, $A = (A^1, A^2, \ldots, A^n)$.

Definiciones. *Sea* A *una matriz de* m $\times$ n, *como la anterior. Cualquiera de las tres operaciones siguientes sobre los renglones* [*columnas*] *de* A *se denomina* operación elemental sobre los renglones [columnas].

(a) *Intercambio de dos renglones* [*columnas*] *cualesquiera de* A.
(b) *Multiplicación de cualquier renglón* [*columna*] *de* A *por una constante no nula.*
(c) *Suma de cualquier múltiplo constante de un renglón* [*columna*] *de* A *a otro renglón* [*columna*].

Cualquiera de las tres operaciones anteriores se denominará operación elemental.

Las operaciones elementales serán tipo 1, tipo 2 *o* tipo 3, *dependiendo de si se trata de (a), (b) o (c).*

Ejemplo 1. Sea

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & -1 & 3 \\ 4 & 0 & 1 & 2 \end{pmatrix}.$$

El intercambio de A_2, el segundo renglón de A, con A_1, el primer renglón de A, es un ejemplo de una operación elemental de tipo 1 sobre los renglones. La matriz resultante es

$$B = \begin{pmatrix} 2 & 1 & -1 & 3 \\ 1 & 2 & 3 & 4 \\ 4 & 0 & 1 & 2 \end{pmatrix}.$$

Así también, la multiplicación de A^2, la segunda columna de A, por 3 es un ejemplo de una operación elemental sobre las columnas del tipo 2. La matriz resultante es

$$C = \begin{pmatrix} 1 & 6 & 3 & 4 \\ 2 & 3 & -1 & 3 \\ 4 & 0 & 1 & 2 \end{pmatrix}.$$

Por último, la suma de A_1, el primer renglón de A, de cuatro veces A_3, el tercer renglón de A, es un ejemplo de una operación elemental con renglones del tipo 3. La matriz resultante es

$$D = \begin{pmatrix} 17 & 2 & 7 & 12 \\ 2 & 1 & -1 & 3 \\ 4 & 0 & 1 & 2 \end{pmatrix}.$$

Definición. *Una* matriz elemental *de $n \times n$ es una matriz obtenida al realizar una operación elemental en* I_n. *Se dice que la matriz es del tipo* 1, 2, *o* 3 *dependiendo de que la operación realizada en* I_n *haya sido del tipo 1, 2 o 3, respectivamente.*

Por ejemplo, el intercambio de los dos primeros renglones de I_3 produce la matriz elemental

$$E = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Nótese que E también se puede obtener mediante el intercambio de las dos primeras columnas de I_3. De hecho, *cualquier matriz elemental puede ser obtenida al menos de dos maneras* —ya sea realizando una operación elemental con renglones en I_n o realizando una operación elemental con columnas en I_n. De igual manera

$$\begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

es una matriz elemental, ya que puede obtenerse a partir de I_3 mediante una operación elemental con columnas del tipo 3 (añadiendo -2 veces la primera columna de I_3 a la tercera) o mediante una operación elemen-

tal con renglones del tipo 3 (añadiendo -2 veces el tercer renglón al primero).

Nuestro primer teorema muestra que realizar una operación elemental en una matriz es equivalente a multiplicar la matriz por una matriz elemental.

Teorema 3.1. *Sea* $A \in M_{m \times n}(F)$, *y supóngase que* B *se obtiene a partir de* A *al realizar una operación elemental con renglones* [*columnas*]. *Entonces existirá una matriz elemental* E *de* $m \times m$ $[n \times n]$ *tal que* $B = EA[B = AE]$. *De hecho,* E *se obtiene al realizar la operación correspondiente con renglones* [*columnas*] *en* $I_m[I_n]$. *Recíprocamente, si* E *es una matriz elemental de* $m \times m[n \times n]$, *entonces* $EA[AE]$ *es una matriz que se puede obtener realizando una operación elemental con renglones* [*columnas*] *en* A.

Antes de considerar una demostración, consideraremos primero un ejemplo para ilustrar el significado del teorema.

Ejemplo 2. Considérese la matriz B del Ejemplo 1. Esta matriz fue obtenida a partir de A (en el Ejemplo 1) intercambiando los dos primeros renglones de A. Realizando esta misma operación en I_3, obtenemos la matriz elemental

$$E = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Nótese que $EA = B$.

En la segunda parte del Ejemplo 1, C se obtiene a partir de A multiplicando la segunda columna de A por 3. Realizando esta misma operación en I_4, obtenemos la matriz elemental

$$E = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Obsérvese que $AE = C$.

DEMOSTRACIÓN DEL TEOREMA 3.1. Supóngase que B se obtiene a partir de A mediante una operación elemental. Debemos considerar seis casos, una por cada tipo de operación con renglones y una por cada tipo de operación con columnas.

Supóngase que B se obtiene intercambiando los renglones p y q de $A (p < q)$ mediante una operación elemental del tipo 1. Entonces

(a) $B_{ij} = A_{ij}$ para $i \neq p$ e $i \neq q$, y $j = 1, 2, \ldots, n$.
(b) $B_{pj} = A_{qj}$ para $j = 1, 2, \ldots, n$.
(c) $B_{qj} = A_{pj}$ para $j = 1, 2, \ldots, n$.

Sea E la matriz elemental de $m \times m$ obtenida a partir de I_m intercambiando los renglones p y q de I_m. Entonces para $i \neq p$ e $i \neq q$ y para toda $j (1 \leq j \leq m)$:

$$\begin{cases} E_{ij} = 0 & \text{si } i \neq j \\ E_{ij} = 1 & \text{si } i = j. \end{cases}$$

Para $i = p$,

$$\begin{cases} E_{pj} = 0 & \text{si } j \neq q \\ E_{pq} = 1. \end{cases}$$

Para $i = q$,

$$\begin{cases} E_{qj} = 0 & \text{si } j \neq p \\ E_{pq} = 1. \end{cases}$$

Puesto que

$$(EA)_{ij} = \sum_{k=1}^{m} E_{ik} A_{kj},$$

para toda j se tiene que

$$\begin{aligned} (EA)_{ij} &= E_{ii} A_{ij} = A_{ij} \qquad \text{si } i \neq p \text{ o } q, \\ (EA)_{pj} &= E_{pq} A_{qj} = A_{qj}, \\ (EA)_{qj} &= E_{qp} A_{pj} = A_{pj}. \end{aligned}$$

De aquí que

$$B_{ij} = (EA)_{ij} \qquad \text{para toda } i \text{ y toda } j.$$

Esto establece el caso 1.

Si B se obtiene a partir de A mediante una operación elemental con renglones del tipo 2 o 3, entonces la demostración es semejante y podrá realizarse como ejercicio.

Ahora supóngase que B se obtiene a partir de A realizando una operación elemental con las columnas de A. Entonces, por el Ejercicio 5, B^t se puede obtener a partir de A^t mediante las correspondientes operaciones elementales con renglones de A. Entonces, las partes anteriores de la demostración muestran que la matriz elemental M de $n \times n$ obtenida al realizar el mismo tipo de operación con renglones en I_n tiene la propiedad de que $B^t = MA^t$. Obsérvese que $E = M^t$ es una matriz elemental que puede obtenerse realizando las operaciones elementales correspondientes con columnas en I_n. Entonces, $B = (B^t)^t = (MA^t)^t = AM^t = AE$, estableciendo el mismo resultado para operaciones con columnas.

La demostración de la proposición recíproca se deja como ejercicio. ■

Es un hecho de gran utilidad el que la inversa de una matriz elemental es también una matriz elemental.

Teorema 3.2. *Las matrices elementales son invertibles y la inversa de una matriz elemental es una matriz elemental del mismo tipo.*

DEMOSTRACIÓN. En vista del hecho de que cualquier matriz elemental de $n \times n$ puede obtenerse mediante operaciones elementales con renglones en I_n, sólo necesitamos considerar tres casos —uno por cada tipo de operación.

Sea E una matriz elemental de $n \times n$.

CASO 1. Supóngase que E se obtiene intercambiando los renglones p y q de $I_n (p \neq q)$, una operación elemental del tipo 1. Resulta fácil verificar que $E^2 = I_n$. Por lo tanto, E es invertible y, de hecho, $E = E^{-1}$. Esto establece el primer caso.

CASO 2. Supóngase que E se obtiene multiplicando el renglón p de I_n por una constante c no nula, una operación elemental del tipo 2. Como $c \neq 0$, c tiene un inverso multiplicativo. Sea $\bar{E}$ la matriz elemental obtenida a partir de I_n multiplicando el renglón p de I_n por c^{-1}. Puede demostrarse fácilmente que $E\bar{E} = \bar{E}E = I_n$. Esto establece el segundo caso.

CASO 3. Supóngase que E se obtiene sumando al renglón p de I_n c veces el renglón q de I_n, donde $p \neq q$ y c es cualquier escalar. Entonces E puede obtenerse a partir de I_n mediante una operación elemental del tipo 3.

Obsérvese que I_n puede obtenerse a partir de E mediante una operación elemental con renglones del tipo 3 —a saber, sumando al renglón p de E $-c$ veces el renglón q de E. Por el Teorema 3.1 se tiene que existe una matriz elemental $\bar{E}$ (del tipo 3) tal que $\bar{E}E = I_n$, y por el Ejercicio 8 de la Sección 2.4 E es invertible y $E^{-1} = \bar{E}$. ■

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.
   (a) Una matriz elemental siempre es cuadrada.
   (b) Los únicos elementos de una matriz elemental son ceros y unos.
   (c) La matriz identidad de $n \times n$ es una matriz elemental.
   (d) El producto de dos matrices elementales de $n \times n$ es una matriz elemental.
   (e) La matriz inversa de una matriz elemental es una matriz elemental.
   (f) La suma de dos matrices elementales de $n \times n$ es una matriz elemental.
   (g) La transpuesta de una matriz elemental es una matriz elemental.
   (h) Si B es una matriz que se puede obtener realizando una operación elemental con renglones en una matriz A, entonces B también puede obtenerse realizando una operación elemental con columnas en A.

(i) Si B es una matriz que se puede obtener realizando una operación elemental con renglones en una matriz A, entonces A se puede obtener realizando una operación elemental con renglones en B.

2. Sean

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & -1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 3 \\ 1 & -2 & 1 \\ 1 & -3 & 1 \end{pmatrix}, \quad \text{y} \quad C = \begin{pmatrix} 1 & 0 & 3 \\ 0 & -2 & -2 \\ 1 & -3 & 1 \end{pmatrix}.$$

Encontrar una operación elemental que permita transformar a A en B y otra que transforme a B en C. Por medio de varias operaciones elementales adicionales transformar a C en I_3.

3. Demostrar la aseveración hecha en la página 141. Una matriz elemental de $n \times n$ puede obtenerse al menos de dos maneras —ya sea realizando una operación elemental con renglones en I_n o realizando una operación elemental con columnas en I_n.

4. Demostrar que E es una matriz elemental si y sólo si E^t lo es.

5. Sea A una matriz de $m \times n$. Demostrar que si B puede obtenerse a partir de A mediante una operación elemental con renglones [columnas], entonces B^t puede obtenerse a partir de A^t mediante la operación elemental correspondiente con renglones [columnas].

6. Completar la demostración del Teorema 3.1.

7. Verificar la aseveración hecha en el Caso 1 de la demostración del Teorema 3.2: Si E es una matriz elemental de $n \times n$ del tipo 1, entonces $E^2 = I_n$.

8. Verificar que para la matriz $\bar{E}$ definida en la demostración del Caso 2 del Teorema 3.2 $E\bar{E} = \bar{E}E = I_n$.

9. Demostrar que cualquier operación elemental con renglones [columnas] del tipo 1 puede obtenerse mediante una sucesión de tres operaciones elementales con renglones [columnas] del tipo 3 seguida por una operación elemental con renglones [columnas] del tipo 2.

10. Demostrar que cualquier operación con renglones [columnas] del tipo 2 puede obtenerse *dividiendo* algún renglón [columna] por un escalar no nulo.

11. Demostrar que cualquier operación elemental con renglones [columnas] del tipo 3 puede obtenerse *restando* un múltiplo de algún renglón [columna] de otro renglón [columna].

3.2 EL RANGO DE UNA MATRIZ Y LA INVERSA DE UNA MATRIZ

En esta sección definiremos el rango de una matriz, y utilizaremos entonces las operaciones elementales para calcular el rango de una matriz o de una transformación lineal. La sección concluirá con un procedimiento para calcular la inversa de una matriz invertible.

Definición. *Si* $A \in M_{m\times n}(F)$, *definimos el* rango *de* A, *que escribiremos* rango(A), *como el rango de la transformación lineal* $L_A\colon F^n \to F^m$.

Un gran número de resultados sobre el rango de las matrices se deriva de inmediato a partir de los hechos correspondientes sobre las transformaciones lineales. Un resultado importante de este tipo, que se deriva del Teorema 2.20 y del Corolario 2 del Teorema 2.21, es que una matriz de $n \times n$ es invertible si y sólo si su rango es n.

Nos gustaría que la definición anterior satisficiera la condición de que el rango de una transformación lineal fuese igual al rango de cualquier matriz que represente dicha transformación. Nuestro primer teorema muestra que, de hecho, esta condición se satisface.

Teorema 3.3. *Sea* $T\colon V \to W$ *una transformación lineal entre espacios vectoriales dimensionalmente finitos y sean* β *y* γ *bases ordenadas para* V *y* W, *respectivamente. Entonces,* $\text{rango}(T) = \text{rango}([T]_\beta^\gamma)$.

DEMOSTRACIÓN. Esta es sólo una manera de reenunciar el Ejercicio 18 de la Sección 2.4. ■

Corolario 1. *Sea* A *una matriz de* $m \times n$. *Si* P *y* Q *son respectivamente matrices invertibles de* $m \times m$ *y* $n \times n$, *entonces rango*(PAQ) = *rango*(A). *En particular rango*(PA) = *rango*(AQ) = *rango*(A).

DEMOSTRACIÓN. Sea $B = PAQ$. En virtud del Ejercicio 12 de la Sección 2.5, existen espacios vectoriales V y W, bases β, β' para V y γ, γ' para W y una transformación lineal $T\colon V \to W$ tal que $A = [T]_\beta^\gamma$ y $B = [T]_{\beta'}^{\gamma'}$. Entonces por el Teorema 3.3

$$\text{rango}(PAQ) = \text{rango}(B) = \text{rango}(T) = \text{rango}(A). \quad ■$$

Corolario 2. *Las operaciones elementales con renglones y columnas en una matriz conservan el rango.*

DEMOSTRACIÓN. Si la matriz B se obtiene a partir de la matriz A mediante una operación elemental con renglones, entonces existe una matriz elemental E tal que $B = EA$. De acuerdo con el Teorema 3.2, E es

invertible y por tanto rango(B) = rango(A) por el Corolario 1. Por lo tanto, las operaciones elementales con renglones conservan el rango. La demostración de que las operaciones elementales con columnas no alteran el rango se deja como ejercicio. ■

El Teorema 3.3 relaciona íntimamente el rango de una transformación lineal con el rango de una matriz. Como las matrices son herramientas útiles para estudiar las transformaciones lineales, es importante desarrollar un método para calcular el rango de una matriz. Esta será nuestra siguiente tarea.

Teorema 3.4. *El rango de cualquier matriz es igual al máximo número de columnas linealmente independientes de dicha matriz; esto es, el rango de una matriz es la dimensión del subespacio generado por las columnas de dicha matriz.*

DEMOSTRACIÓN. Para toda $A \in \mathsf{M}_{m\times n}(F)$,

$$\text{rango}(A) = \text{rango}(\mathsf{L}_A) = \dim(\mathsf{R}(\mathsf{L}_A)).$$

Sea $\beta = \{e_1, e_2, \ldots, e_n\}$, la base ordenada estándar para F^n. Entonces β genera a F^n y entonces

$$\mathsf{R}(\mathsf{L}_A) = L\{\mathsf{L}_A(e_1), \mathsf{L}_A(e_2), \ldots, \mathsf{L}_A(e_n)\}.$$

Pero hemos visto que $\mathsf{L}_A(e_j) = A^j$, la columna j-ésima de A. Por lo tanto

$$\mathsf{R}(\mathsf{L}_A) = L\{A^1, A^2, \ldots, A^n\}.$$

Entonces

$$\text{rango}(A) = \dim(\mathsf{R}(\mathsf{L}_A)) = \dim(L\{A^1, A^2, \ldots, A^n\}). \quad ■$$

Ejemplo 3. Sea

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

Obsérvese que la primera y la segunda columnas de A son linealmente independientes y que la tercera columna es una combinación lineal de las dos primeras. Entonces

$$\text{rango}(A) = \dim\left(L\left\{\begin{pmatrix}1\\0\\1\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}\right\}\right) = 2.$$

Para calcular el rango de una matriz A, a menudo es útil posponer el empleo del Teorema 3.4 hasta que A se haya modificado adecuadamente por medio de operaciones elementales con renglones y columnas de

tal manera que el número de columnas linealmente independientes sea evidente. El Corolario 2 del Teorema 3.3 garantiza que el rango de la matriz modificada es el mismo que el rango de A. Una de estas modificaciones de A se puede obtener mediante el uso de operaciones elementales con columnas y renglones hasta introducir elementos nulos. El ejemplo siguiente ilustra este procedimiento.

Ejemplo 4. Sea

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{pmatrix}.$$

Si restamos el primer renglón de A de los renglones 2 y 3 (operaciones elementales tipo 3 con renglones), se tiene como resultado

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & -2 & 2 \\ 0 & -1 & 1 \end{pmatrix}.$$

Si ahora restamos dos veces la primera columna de la segunda y la primera columna de la tercera (operaciones elementales tipo 3 con columnas) obtenemos

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 2 \\ 0 & -1 & 1 \end{pmatrix}.$$

Es ahora evidente que el número máximo de columnas linealmente independientes de esta matriz es 2. Por lo tanto, el rango de A es 2.

El siguiente teorema utiliza este proceso de modificación de una matriz por medio de operaciones elementales con renglones y columnas para transformarla a una forma particularmente simple. La fuerza de este teorema se puede ver en sus corolarios.

Teorema 3.5. *Sea* A *una matriz de* $m \times n$ *de rango* r. *Entonces* $r \leq m$, $r \leq n$, *y por medio de un número finito de operaciones elementales con renglones y columnas* A *se puede transformar en una matriz* D *tal que*

(a) $D_{ij} = 0$ *para* $i \neq j$,
(b) $D_{ii} = 1$ *para* $i \leq r$,
(c) $D_{ii} = 0$ *para* $i > r$.

El teorema anterior y sus corolarios son muy importantes. Su demostración, aunque fácil de entender, es bastante tediosa. Como ayuda para seguir la demostración consideraremos primero un ejemplo.

Ejemplo 5. Considérese la matriz

$$A = \begin{pmatrix} 0 & 2 & 4 & 2 & 2 \\ 4 & 4 & 4 & 8 & 0 \\ 8 & 2 & 0 & 10 & 2 \\ 6 & 3 & 2 & 9 & 1 \end{pmatrix}.$$

Por medio de una sucesión de operaciones elementales con renglones y columnas transformaremos a A en una matriz D tal como lo establece el Teorema 3.5. Escribiremos muchas de las matrices intermedias, pero en algunas ocasiones transformaremos una matriz a partir de la anterior mediante varias operaciones elementales simultáneas. El número sobre la flecha indicará cuántas operaciones se han involucrado; trátese de identificar la naturaleza de cada operación (si se hizo sobre un renglón o sobre una columna y el tipo de operación).

$$\begin{pmatrix} 0 & 2 & 4 & 2 & 2 \\ 4 & 4 & 4 & 8 & 0 \\ 8 & 2 & 0 & 10 & 2 \\ 6 & 3 & 2 & 9 & 1 \end{pmatrix} \xrightarrow{1} \begin{pmatrix} 4 & 4 & 4 & 8 & 0 \\ 0 & 2 & 4 & 2 & 2 \\ 8 & 2 & 0 & 10 & 2 \\ 6 & 3 & 2 & 9 & 1 \end{pmatrix} \xrightarrow{1} \begin{pmatrix} 1 & 1 & 1 & 2 & 0 \\ 0 & 2 & 4 & 2 & 2 \\ 8 & 2 & 0 & 10 & 2 \\ 6 & 3 & 2 & 9 & 1 \end{pmatrix} \xrightarrow{2}$$

$$\begin{pmatrix} 1 & 1 & 1 & 2 & 0 \\ 0 & 2 & 4 & 2 & 2 \\ 0 & -6 & -8 & -6 & 2 \\ 0 & -3 & -4 & -3 & 1 \end{pmatrix} \xrightarrow{3} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 4 & 2 & 2 \\ 0 & -6 & -8 & -6 & 2 \\ 0 & -3 & -4 & -3 & 1 \end{pmatrix} \xrightarrow{1}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 1 \\ 0 & -6 & -8 & -6 & 2 \\ 0 & -3 & -4 & -3 & 1 \end{pmatrix} \xrightarrow{2} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 1 & 1 \\ 0 & 0 & 4 & 0 & 8 \\ 0 & 0 & 2 & 0 & 4 \end{pmatrix} \xrightarrow{3} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 8 \\ 0 & 0 & 2 & 0 & 4 \end{pmatrix} \xrightarrow{1}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 & 4 \end{pmatrix} \xrightarrow{1} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{1} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} = D.$$

En virtud del Corolario 2 del Teorema 3.3, rango(A) = rango(D), pero es evidente que rango$(D) = 3$; así rango$(A) = 3$. Nótese que las dos primeras operaciones elementales dan como resultado un 1 en la posición 1, 1 y las siguientes operaciones (tipo 3) dan como resultado ceros, en todas las posiciones del primer renglón y primera columna a excepción de la posición 1, 1. Las operaciones elementales subsecuentes no producen cambios en el primer renglón ni en la primera columna. Con este ejemplo en mente procederemos a la demostración.

DEMOSTRACIÓN DEL TEOREMA 3.5. Si A es una matriz nula, $r = 0$ por el Ejercicio 3. En este caso se concluye que $D = A$.

Ahora, supóngase que $A \neq O$ y $r = \text{rango}(A)$; entonces $r > 0$. La demostración se hará por inducción matemática sobre m, el número de renglones de A.

Supóngase que $m = 1$. Por medio de a lo más una operación con columnas del tipo 1 y una operación con columnas del tipo 2, A puede ser transformada en una matriz con un 1 en la posición 1, 1, y por medio de a lo más $n - 1$ operaciones con columnas del tipo 3, esta matriz puede ser transformada a su vez en la matriz

$$D = (1, 0, 0, \ldots, 0).$$

Nótese que existe un máximo de una columna linealmente independiente en D. Por lo tanto $\text{rango}(D) = \text{rango}(A) = 1$ de acuerdo con el Corolario 2 del Teorema 3.3 y el Teorema 3.4. Por lo tanto, el teorema queda establecido para $m = 1$.

Ahora supóngase que el teorema se cumple para cualquier matriz con un máximo de $m - 1$ renglones (para algún $m > 1$). Demostraremos que el teorema se cumple para cualquier matriz con m renglones.

Supóngase que A es cualquier matriz de $m \times n$. Si $n = 1$, el Teorema 3.5 se demuestra de una manera análoga que para $m = 1$. (Véase el Ejercicio 10.)

Supondremos que $n > 1$. Como $A \neq O$, $A_{ij} \neq 0$ para algún i, j. Por medio de a lo más una operación con renglones y una operación con columnas (ambas del tipo 1) se puede colocar un elemento no nulo en la posición 1, 1 (tal como se hizo en el Ejemplo 5). Por medio de a lo más una operación adicional del tipo 2 podemos asegurar un 1 en la posición 1, 1. (Véase la segunda operación en el Ejemplo 5.) Por medio de un máximo de $m - 1$ operaciones con renglones del tipo 3 y $n - 1$ operaciones con columnas del tipo 3 podemos eliminar a todos los elementos no nulos del primer renglón y la primera columna, con excepción del 1 en la posición 1, 1. (En el Ejemplo 5 utilizamos dos operaciones con renglones y tres operaciones con columnas para poder alcanzar este estado.)

Por tanto, con un número finito de operaciones elementales, A puede transformarse en una matriz

$$B = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \cdot & & & \\ \cdot & & B' & \\ \cdot & & & \\ 0 & & & \end{array}\right),$$

donde B' es una matriz de $(m-1)\times(n-1)$. En el Ejemplo 5,

$$B' = \begin{pmatrix} 2 & 4 & 2 & 2 \\ -6 & -8 & -6 & 2 \\ -3 & -4 & -3 & 1 \end{pmatrix}.$$

Por el Ejercicio 11, B' tiene un rango una unidad menor que el de B. Como rango(A) = rango$(B) = r$, rango$(B') = r-1$. De acuerdo con la hipótesis de inducción $r-1 \leq n-1$ y $r-1 \leq m-1$. Por lo tanto, $r \leq m$ y $r \leq n$.

También por la hipótesis de inducción, B' puede transformarse por medio de un número finito de operaciones elementales con renglones y columnas en una matriz D' de $(m-1)\times(n-1)$ tal que

$$\begin{aligned} (D')_{i,j} &= 0 \qquad \text{si } i \neq j, \\ (D')_{i,i} &= 1 \qquad \text{si } i \leq r-1, \\ (D')_{i,i} &= 0 \qquad \text{si } i \geq r. \end{aligned}$$

Esto es, que D' consta totalmente de ceros a excepción de las primeras $r-1$ posiciones de la diagonal principal que tienen unos. Sea

$$D = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \cdot & & & \\ \cdot & & D' & \\ \cdot & & & \\ 0 & & & \end{array}\right).$$

Vemos que el teorema se establece una vez que queda demostrado que D puede obtenerse de B por medio de un número finito de operaciones elementales con renglones y columnas. Pero esto se obtiene aplicando repetidas veces el Ejercicio 12.

Así, como A puede ser transformada en B y B puede ser transformada en D, ambas mediante un número finito de operaciones elementales, entonces A puede ser transformada en D mediante un número finito de operaciones elementales.

Finalmente, como D' contiene unos en las primeras $r-1$ posiciones sobre la diagonal principal, D contiene unos en las primeras r posiciones sobre su diagonal principal y ceros en el resto de ellas. Luego entonces, $D_{ii} = 1$ si $i \leq r$, $D_{ii} = 0$ si $i > r$, y $D_{ij} = 0$ si $i \neq j$. Esto establece el teorema. ■

Corolario 1. *Sea* A *una matriz de* $m \times n$ *de rango* r. *Entonces existen matrices invertibles* B *y* C *de dimensiones* $m \times m$ *y* $n \times n$, *respectivamente, tales que* D = BAC *donde* D *es una matriz de* $m \times n$ *que satisface*

(a) $D_{ij} = 0$ si $i \neq j$,
(b) $D_{ii} = 1$ si $i \leq r$,
(c) $D_{ii} = 0$ si $i > r$.

DEMOSTRACIÓN. De acuerdo con el Teorema 3.5, A puede ser transformada en la matriz D mediante un número finito de operaciones elementales con renglones y columnas. Podemos recurrir al Teorema 3.1 cada vez que realicemos operaciones elementales. Entonces, existirán matrices elementales de $m \times m$ $E_1, E_2, \ldots, E_p$ y matrices elementales de $n \times n$ $G_1, G_2, \ldots, G_q$ tales que

$$D = E_pE_{p-1} \ldots E_2E_1AG_1G_2 \ldots G_q.$$

De acuerdo con el Teorema 3.2, todas las E_j y G_j son invertibles. Sean $B = E_pE_{p-1} \ldots E_1$ y $C = G_1 \ldots G_q$. Entonces, de acuerdo con el Ejercicio 2 de la Sección 2.4 B y C son invertibles y $D = BAC$. ■

Corolario 2. *Sea* A *una matriz arbitraria de* $m \times n$.

(a) *Rango*(A^t) = *rango*(A).
(b) *El rango de cualquier matriz es igual al número máximo de renglones linealmente independientes de dicha matriz; esto es, el rango de una matriz es la dimensión del subespacio generado por los renglones de la matriz.*
(c) *Los renglones y las columnas de cualquier matriz generan subespacios de la misma dimensión, numéricamente iguales al rango de la matriz.*

DEMOSTRACIÓN.

(a) Por el Corolario 1 existen matrices invertibles B y C tales que $D = BAC$, donde D satisface las condiciones enunciadas en el corolario. Tomando las transpuestas tenemos

$$D^t = C^tA^tB^t.$$

Puesto que B y C son invertibles, también lo son B^t y C^t, en virtud del Ejercicio 3 de la Sección 2.4. Entonces, por el Corolario 1 del Teorema 3.3,

$$\text{rango}(A^t) = \text{rango}(C^tA^tB^t) = \text{rango}(D^t).$$

Supóngase que $r = \text{rango}(A)$. Entonces D^t es una matriz de $n \times m$ que satisface las condiciones del Corolario 1 y, por tanto, $\text{rango}(D^t) = r$ por el Teorema 3.4. Así,

$$\text{rango}(A^t) = \text{rango}(D^t) = r = \text{rango}(A).$$

Esto establece (a).

Las demostraciones de (b) y (c) se dejan como ejercicios. ■

Corolario 3. *Cualquier matriz invertible es un producto de matrices elementales.*

DEMOSTRACIÓN. Si A es una matriz invertible de $n \times n$, entonces rango$(A) = n$, y entonces por el Corolario 1 existen matrices invertibles B y C tales que $D = BAC$, donde $D_{ij} = 0$ para $i \neq j$ y $D_{ii} = 1$ para $1 \leq i \leq n$. Entonces $D = I_n$; esto es, $I_n = BAC$.

En la demostración del Corolario 1 nótese que $B = E_p E_{p-1} \ldots E_1$ y $C = G_1 G_2 \ldots G_q$, donde las E_i y las G_i son matrices elementales. Entonces $A = B^{-1} I_n C^{-1} = B^{-1} C^{-1}$, y así $A = E_1^{-1} E_2^{-1} \ldots E_p^{-1} G_q^{-1} G_{q-1}^{-1} \ldots G_1^{-1}$. Pero la inversa de una matriz elemental es elemental y por lo tanto A es el producto de matrices elementales. ■

Utilizaremos al Corolario 2 para relacionar el rango de un producto de matrices con el rango de cada factor. Nótese cómo la demostración emplea la relación entre el rango de una matriz y el rango de una transformación lineal.

Teorema 3.6. *Sean* $\mathsf{T}: \mathsf{V} \to \mathsf{W}$ *y* $\mathsf{U}: \mathsf{W} \to \mathsf{Z}$ *transformaciones lineales en espacios vectoriales dimensionalmente finitos* V, W *y* Z, *y sean* A *y* B *matrices tales que* AB *está definido. Entonces*

(a) *rango*(UT) $\leq$ *rango*(U) .

(b) *rango*(UT) $\leq$ *rango*(T) .

(c) *rango*(AB) $\leq$ *rango*(A) .

(d) *rango*(AB) $\leq$ *rango*(B) .

DEMOSTRACIÓN. Claramente se ve que $\mathsf{R}(\mathsf{T}) \subseteq \mathsf{W}$ y por lo tanto

$$\mathsf{R}(\mathsf{UT}) = \mathsf{UT}(\mathsf{V}) = \mathsf{U}(\mathsf{R}(\mathsf{T})) \subseteq \mathsf{U}(\mathsf{W}) = \mathsf{R}(\mathsf{U}).$$

Entonces

$$\text{rango}(\mathsf{UT}) = \dim(\mathsf{R}(\mathsf{UT})) \leq \dim(\mathsf{R}(\mathsf{U})) = \text{rango}(\mathsf{U}).$$

Esto establece el inciso (a).

De acuerdo con el inciso (a)

$$\text{rango}(AB) = \text{rango}(\mathsf{L}_{AB}) = \text{rango}(\mathsf{L}_A \mathsf{L}_B) \leq \text{rango}(\mathsf{L}_A) = \text{rango}(A).$$

Esto establece el inciso (c).

En virtud del inciso (c) y del Corolario 2 del Teorema 3.5

$$\text{rango}(AB) = \text{rango}((AB)^t) = \text{rango}(B^t A^t) \leq \text{rango}(B^t) = \text{rango}(B).$$

Esto establece el inciso (d).

Sean α, β y γ bases ordenadas para V, W y Z, respectivamente, y sean $A' = [\mathsf{U}]_\beta^\gamma$ y $B' = [\mathsf{T}]_\alpha^\beta$. Entonces $A'B' = [\mathsf{UT}]_\alpha^\gamma$ de acuerdo con el Teorema 2.12. Por lo tanto, por el Teorema 3.3 y el inciso (d),

$$\text{rango}(\mathsf{UT}) = \text{rango}(A'B') \leq \text{rango}(B') = \text{rango}(\mathsf{T}).$$

Esto establece el inciso (b). ■

Veremos posteriormente que es importante poder calcular el rango de cualquier matriz. Podemos emplear el Corolario 2 del Teorema 3.3, los Teoremas 3.4 y 3.5 y el Corolario 2 del Teorema 3.5 para llevar a cabo nuestro propósito.

El objetivo es utilizar operaciones elementales con renglones y columnas en una matriz para "simplificarla" (de tal modo que la matriz transformada tenga muchos ceros) hasta que una observación sencilla nos permita determinar cuántos renglones o columnas linealmente independientes tiene la matriz y así determinar su rango.

Ejemplo 6.

(a) Sea

$$A = \begin{pmatrix} 1 & 2 & 1 & 3 \\ 1 & 1 & -1 & 1 \end{pmatrix}.$$

Nótese que los renglones de A son linealmente independientes pues uno no es múltiplo del otro. Luego, rango$(A) = 2$.

(b) Sea

$$A = \begin{pmatrix} 1 & 3 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 3 & 0 & 0 \end{pmatrix}.$$

En este caso hay varias maneras para proceder. Supóngase que principiamos con una operación elemental con renglones para obtener un cero en la posición 2, 1. Restando el primer renglón del segundo, obtenemos

$$\begin{pmatrix} 1 & 3 & 1 & 1 \\ 0 & -3 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}.$$

Ahora nótese que el tercer renglón es un múltiplo del segundo, y que el primero y el segundo renglones son linealmente independientes. Por tanto, rango$(A) = 2$.

Como un método alternativo, nótese que la primera, la tercera y la cuarta columnas de A son idénticas y que la primera y la segunda columnas de A son linealmente independientes. Por lo tanto, rango$(A) = 2$.

(c) Sea

$$A = \begin{pmatrix} 1 & 2 & 3 & 1 \\ 2 & 1 & 1 & 1 \\ 1 & -1 & 1 & 0 \end{pmatrix}.$$

Utilizando diversas operaciones elementales con renglones y columnas, obtenemos la siguiente secuencia de matrices:

$$\begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & -3 & -5 & -1 \\ 0 & -3 & -2 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & -3 & -5 & 1 \\ 0 & 0 & 3 & 0 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -3 & -5 & 1 \\ 0 & 0 & 3 & 0 \end{pmatrix}, \quad \text{y} \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{pmatrix}.$$

Es claro que la última matriz tiene tres renglones linealmente independientes y, por lo tanto, su rango es 3.

En síntesis, realícense operaciones en renglones y columnas hasta que la matriz se haya simplificado lo suficiente, de tal manera que el máximo número de renglones o columnas linealmente independientes se haga evidente.

La Inversa de una Matriz

Hemos afirmado que una matriz de $n \times n$ es invertible si y sólo si su rango es n. Como ya sabemos cómo calcular el rango de cualquier matriz, ya podemos determinar si una matriz es o no invertible. Proporcionaremos ahora una técnica sencilla para calcular la inversa de una matriz, la cual empleará operaciones elementales con renglones.

Definición. *Sean* A *y* B *matrices de* $\mathrm{m} \times \mathrm{n}$ *y de* $\mathrm{m} \times \mathrm{p}$, *respectivamente. Por* matriz aumentada $(\mathrm{A} \mid \mathrm{B})$ *entenderemos la matriz de* $\mathrm{m} \times (\mathrm{n} + \mathrm{p})$

$$(\mathrm{A}^1, \ldots, \mathrm{A}^n, \mathrm{B}^1, \ldots, \mathrm{B}^p),$$

donde A^i *y* B^j *significan la* i*-ésima columna de* A *y la* j*-ésima columna de* B, *respectivamente.*

Sea A una matriz invertible de $n \times n$ y considérese la matriz aumentada $C = (A \mid I_n)$ de $n \times 2n$. En virtud del Ejercicio 15 tenemos

$$A^{-1}C = (A^{-1}A \mid A^{-1}I_n) = (I_n \mid A^{-1}). \tag{1}$$

Por el Corolario 3 del Teorema 3.5, A^{-1} es el producto de matrices elementales, digamos $A^{-1} = E_pE_{p-1} \ldots E_1$. Entonces la Ecuación (1) se transforma en

$$E_pE_{p-1} \ldots E_1(A \mid I_n) = A^{-1}C = (I_n \mid A^{-1}).$$

Como multiplicar a una matriz por la izquierda por una matriz elemental transforma a la matriz en la misma forma que una operación elemental con renglones (Teorema 3.1), tenemos el siguiente resultado: Si A es una matriz invertible de $n \times n$, entonces es posible transformar la matriz $(A \mid I_n)$ en la matriz $(I_n \mid A^{-1})$ por medio de un número finito de operaciones elementales *con renglones*.

Recíprocamente, supóngase que A es invertible y que la matriz $(A \mid I_n)$ puede ser transformada en la matriz $(I_n \mid B)$ mediante un número finito de operaciones elementales con renglones. Sean $E_1, E_2, \ldots, E_p$ las matrices elementales asociadas con estas operaciones elementales con renglones como en el Teorema 3.1; entonces

$$E_p \ldots E_2E_1(A \mid I_n) = (I_n \mid B). \tag{2}$$

Haciendo $M = E_p \ldots E_2E_1$, tenemos de la ecuación (2) que

$$M(A \mid I_n) = (MA \mid M) = (I_n \mid B).$$

Por lo tanto $MA = I_n$ y $M = B$, y se tiene entonces que $M = A^{-1}$. Así $B = M = A^{-1}$. Así se tiene el siguiente resultado: *Si* A *es una matriz invertible de* n $\times$ n *y si la matriz* (A | I$_n$) *se transforma en una matriz de la forma* (I$_n$ | B) *por medio de un número finito de operaciones elementales con renglones, entonces* B = A^{-1}.

El ejemplo siguiente ilustra este procedimiento.

Ejemplo 7. Calcularemos la inversa de la matriz

$$A = \begin{pmatrix} 0 & 2 & 4 \\ 2 & 4 & 2 \\ 3 & 3 & 1 \end{pmatrix}.$$

(El lector puede verificar que rango$(A) = 3$ para estar seguro de que A es invertible.) Para calcular A^{-1} debemos utilizar operaciones elementales con renglones para transformar la matriz

$$(A \mid I) = \begin{pmatrix} 0 & 2 & 4 & 1 & 0 & 0 \\ 2 & 4 & 2 & 0 & 1 & 0 \\ 3 & 3 & 1 & 0 & 0 & 1 \end{pmatrix}$$

en $(I \mid A^{-1})$. El método más eficiente para llevar a cabo esta transformación es transformar sucesivamente las columnas de A, empezando por la primera columna, en la correspondiente columna de I. Como necesitamos

un elemento no nulo en la posición 1,1, principiaremos intercambiando los renglones 1 y 2. El resultado es

$$\begin{pmatrix} 2 & 4 & 2 & 0 & 1 & 0 \\ 0 & 2 & 4 & 1 & 0 & 0 \\ 3 & 3 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

A fin de colocar un 1 en la posición 1,1 debemos multiplicar el primer renglón por $\frac{1}{2}$; esta operación nos da

$$\begin{pmatrix} 1 & 2 & 1 & 0 & \frac{1}{2} & 0 \\ 0 & 2 & 4 & 1 & 0 & 0 \\ 3 & 3 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

Ahora completamos nuestra labor sobre la primera columna al sumar al tercer renglón -3 veces el primero para obtener

$$\begin{pmatrix} 1 & 2 & 1 & 0 & \frac{1}{2} & 0 \\ 0 & 2 & 4 & 1 & 0 & 0 \\ 0 & -3 & -2 & 0 & -\frac{3}{2} & 1 \end{pmatrix}.$$

Para transformar la segunda columna de la matriz anterior en la segunda columna de I multiplicaremos el renglón 2 por $\frac{1}{2}$ para obtener un 1 en la posición 2,2. Esta operación da

$$\begin{pmatrix} 1 & 2 & 1 & 0 & \frac{1}{2} & 0 \\ 0 & 1 & 2 & \frac{1}{2} & 0 & 0 \\ 0 & -3 & -2 & 0 & -\frac{3}{2} & 1 \end{pmatrix}.$$

Podemos ahora completar el trabajo en la segunda columna sumando -2 veces el renglón 2 al renglón 1 y 3 veces el renglón 2 al renglón 3. El resultado es

$$\begin{pmatrix} 1 & 0 & -3 & -1 & \frac{1}{2} & 0 \\ 0 & 1 & 2 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 4 & \frac{3}{2} & -\frac{3}{2} & 1 \end{pmatrix}.$$

Unicamente queda por transformar la tercera columna. Para colocar un 1 en la posición 3,3 multiplicamos el renglón 3 por $\frac{1}{4}$; esta operación da

$$\begin{pmatrix} 1 & 0 & -3 & -1 & \frac{1}{2} & 0 \\ 0 & 1 & 2 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & \frac{3}{8} & -\frac{3}{8} & \frac{1}{4} \end{pmatrix}.$$

La suma de múltiplos apropiados del renglón 3 a los renglones 1 y 2 completa el proceso y da

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{8} & -\frac{5}{8} & \frac{3}{4} \\ 0 & 1 & 0 & -\frac{1}{4} & \frac{3}{4} & -\frac{1}{2} \\ 0 & 0 & 1 & \frac{3}{8} & -\frac{3}{8} & \frac{1}{4} \end{pmatrix}.$$

Entonces

$$A^{-1} = \begin{pmatrix} \frac{1}{8} & -\frac{5}{8} & \frac{3}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{2} \\ \frac{3}{8} & -\frac{3}{8} & \frac{1}{4} \end{pmatrix}.$$

El ser capaces de calcular la inversa de una matriz nos permite calcular el inverso de una transformación lineal. El ejemplo siguiente muestra la técnica.

Ejemplo 8. Sea T: $\mathsf{P}_2(R) \rightarrow \mathsf{P}_2(R)$ definida mediante $\mathsf{T}(f) = f + f' + f''$, donde f' y f'' son la primera y la segunda derivadas de f. Se puede demostrar fácilmente que $\mathsf{N}(\mathsf{T}) = \{0\}$, de manera que T es invertible. Tomando $\beta = \{1, x, x^2\}$, tenemos

$$[\mathsf{T}]_\beta = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Y encontramos que la inversa de la matriz es

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Pero $([\mathsf{T}]_\beta)^{-1} = [\mathsf{T}^{-1}]_\beta$ por el Corolario 1 del Teorema 2.21. Por lo tanto, por el Teorema 2.16, tenemos que

$$\mathsf{T}^{-1}(a_0 + a_1x + a_2x^2) = (a_0 - a_1) + (a_1 - 2a_2)x + a_2x^2.$$

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

   (a) El rango de una matriz es igual al número de sus columnas no nulas.
   (b) El producto de dos matrices siempre tiene rango igual al menor de los rangos de las dos matrices.
   (c) La matriz cero de $m \times n$ es la única matriz de $m \times n$ de rango 0.
   (d) Las operaciones elementales con renglones conservan el rango.
   (e) Las operaciones elementales con columnas no necesariamente conservan el rango.

(f) El rango de una matriz es igual al máximo número de renglones linealmente independientes de la matriz.
(g) La inversa de una matriz se puede calcular exclusivamente por medio de operaciones elementales con renglones.
(h) El rango de una matriz de $n \times n$ es a lo más n.
(i) Una matriz de $n \times n$ de rango n es invertible.

2. Encontrar el rango de las matrices siguientes:

(a) $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ (b) $\begin{pmatrix} 1 & 1 & 0 \\ 2 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ (c) $\begin{pmatrix} 1 & 0 & 2 \\ 1 & 1 & 4 \end{pmatrix}$

(d) $\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \end{pmatrix}$ (e) $\begin{pmatrix} 1 & 2 & 3 & 1 & 1 \\ 1 & 4 & 0 & 1 & 2 \\ 0 & 2 & -3 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$

(f) $\begin{pmatrix} 1 & 2 & 0 & 1 & 1 \\ 2 & 4 & 1 & 3 & 0 \\ 3 & 6 & 2 & 5 & 1 \\ -4 & -8 & 1 & -3 & 1 \end{pmatrix}$ (g) $\begin{pmatrix} 1 & 1 & 0 & 1 \\ 2 & 2 & 0 & 2 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}$

3. Demostrar que para cualquiera matriz A de $m \times n$, $\text{rango}(A) = 0$ si y sólo si A es la matriz nula.

4. Utilizar operaciones elementales con renglones y columnas para transformar cada una de las matrices siguientes en una matriz D que satisfaga las condiciones del Teorema 3.5, y luego determinar el rango de cada matriz.

(a) $\begin{pmatrix} 1 & 1 & 1 & 2 \\ 2 & 0 & -1 & 2 \\ 1 & 1 & 1 & 2 \end{pmatrix}$ (b) $\begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 2 & 1 \end{pmatrix}$

5. Para cada una de las siguientes matrices calcular el rango y la inversa, si ésta existe.

(a) $\begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}$ (b) $\begin{pmatrix} 0 & -2 & 4 \\ 1 & 1 & -1 \\ 2 & 4 & -5 \end{pmatrix}$ (c) $\begin{pmatrix} 1 & 2 & 1 \\ -1 & 1 & 2 \\ 1 & 0 & 1 \end{pmatrix}$

(d) $\begin{pmatrix} 1 & 2 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ (e) $\begin{pmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & -1 & 2 \\ 2 & 0 & 1 & 0 \\ 0 & -1 & 1 & -3 \end{pmatrix}$

6. Para cada una de las siguientes transformaciones lineales T, determinar si T es invertible y calcular T^{-1} si existe.

(a) T: $P_2(R) \to P_2(R)$ definida por $T(f) = f'' + 2f' - f$

(b) T: $R^3 \to R^3$ definida por

$$T(a_1, a_2, a_3) = (a_1 + 2a_2 + a_3, -a_1 + a_2 + 2a_3, a_1 + a_3)$$

(c) T: $R^3 \to P_2(R)$ definida por

$$T(a_1, a_2, a_3) = (a_1 + a_2 + a_3) + (a_1 - a_2 + a_3)x + a_1x^2$$

7. Expresar la matriz invertible

$$\begin{pmatrix} 1 & 2 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

como un producto de matrices elementales.

8. Sea A una matriz de $m \times n$. Demostrar que si c es cualquier escalar no nulo, entonces rango(cA) = rango(A).

9. Completar la demostración del Corolario 2 del Teorema 3.3 demostrando que las operaciones elementales con columnas conservan el rango.

10. Demostrar el Teorema 3.5 para el caso en que A es una matriz de $m \times 1$.

11. Sea

$$B = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & B' & \\ 0 & & & \end{array}\right),$$

donde B' es una submatriz de $m \times n$. Demostrar que si rango$(B) = r$, entonces rango$(B') = r - 1$.

12. Sean B' y D' matrices de $m \times n$ y sean B y D las matrices de $(m + 1) \times (n + 1)$ definidas por

$$B = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & B' & \\ 0 & & & \end{array}\right) \quad \text{y} \quad D = \left(\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & D' & \\ 0 & & & \end{array}\right).$$

Demostrar que si B' puede transformarse en D' mediante una operación elemental con renglones [columnas], entonces B puede transformarse en D mediante una operación elemental con renglones.

13. Demostrar los incisos (b) y (c) del Corolario 2 del Teorema 3.5.

14. Sean T, U: $\mathsf{V} \to \mathsf{W}$ transformaciones lineales. Demostrar que

(a) $\mathsf{R}(\mathsf{T}+\mathsf{U}) \subseteq \mathsf{R}(\mathsf{T}) + \mathsf{R}(\mathsf{U})$.
(b) Si W es dimensionalmente finito, entonces $\text{rango}(\mathsf{T}+\mathsf{U}) \leq \text{rango}(\mathsf{T}) + \text{rango}(\mathsf{U})$.
(c) Deducir del inciso (b) que, para cualquier par de matrices A y B de $m \times n$, $\text{rango}(A+B) \leq \text{rango}(A) + \text{rango}(B)$.

15. Si A y B son matrices de n renglones, demostrar que $M(A \mid B) = (MA \mid MB)$ para cualquier matriz M de $m \times n$.

16. Demostrar que si B es una matriz de 3×1 y C una matriz de 1×3 entonces el rango de la matriz BC de 3×3 es a lo más 1. Recíprocamente, demostrar que si A es cualquier matriz de 3×3 con rango 1, entonces existen una matriz B de 3×1 y una matriz C de 1×3 tales que $A = BC$.

3.3 SISTEMAS DE ECUACIONES LINEALES: ASPECTOS TEORICOS

Esta sección y la siguiente están dedicadas al estudio de los sistemas de ecuaciones lineales, los cuales se presentan de manera natural tanto en las ciencias físicas como en las sociales. En esta sección aplicaremos los resultados del Capítulo 2 para describir a los conjuntos solución de los sistemas de ecuaciones lineales como subconjuntos de un espacio vectorial. En la Sección 3.4 se utilizarán operaciones elementales con renglones para proporcionar un método de cálculo para encontrar todas las soluciones a tales sistemas.

El sistema de ecuaciones

$$(S)\quad \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m, \end{cases}$$

donde a_{ij} y $b_i (1 \leq i \leq m$ y $1 \leq j \leq n)$ son elementos de un campo F y $x_1, x_2, \ldots, x_n$ son n variables que toman valores en F, se denomina un *sistema de* m *ecuaciones lineales con* n *incógnitas sobre el campo* F.

La matriz de $m \times n$

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

se llama la *matriz de coeficientes* del sistema (S).

Si expresamos a X y B como

$$X = \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} \quad \text{y} \quad B = \begin{pmatrix} b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{pmatrix},$$

entonces el sistema (S) se puede reescribir como una ecuación matricial única

$$AX = B.$$

Para utilizar los resultados que hemos desarrollado hasta ahora, a menudo consideraremos a un sistema de ecuaciones lineales como una ecuación matricial única.

Una *solución* del sistema (S) es una n-dimensional

$$s = \begin{pmatrix} s_1 \\ \cdot \\ \cdot \\ \cdot \\ s_n \end{pmatrix} \in \mathsf{F}^n$$

tal que $As = B$. El conjunto de todas las soluciones del sistema (S) se llama *conjunto solución* del sistema.

Ejemplo 9.

(a) Considérese el sistema

$$\begin{cases} x_1 + x_2 = 3 \\ x_1 - x_2 = 1. \end{cases}$$

Utilizando técnicas conocidas podemos resolver el sistema anterior y concluir que existe una solución única: $x_1 = 2$, $x_2 = 1$; es decir,

$$s = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

El sistema puede escribirse de forma matricial como

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix};$$

entonces

$$A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \text{ y } B = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

(b) Considérese

$$\begin{cases} 2x_1 + 3x_2 + x_3 = 1 \\ x_1 - x_2 + 2x_3 = 6; \end{cases}$$

o bien,

$$\begin{pmatrix} 2 & 3 & 1 \\ 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 6 \end{pmatrix}.$$

Este sistema tiene muchas soluciones tales como

$$s = \begin{pmatrix} -6 \\ 2 \\ 7 \end{pmatrix} \quad \text{y} \quad s = \begin{pmatrix} 8 \\ -4 \\ -3 \end{pmatrix}.$$

(c) Considérese

$$\begin{cases} x_1 + x_2 = 0 \\ x_1 + x_2 = 1; \end{cases}$$

o bien,

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Es evidente que este sistema no tiene soluciones y entonces vemos que un sistema de ecuaciones lineales puede tener una, muchas o ninguna solución.

Debemos ser capaces de reconocer cuándo un sistema tiene soluciones y luego ser capaces de describir todas las soluciones. Esta sección y la siguiente estarán dedicadas a este fin.

Principiaremos nuestro estudio de sistemas de ecuaciones examinando el tipo de sistemas "homogéneos" de ecuaciones lineales. Como veremos posteriormente (Teorema 3.7), el conjunto de soluciones de un sistema homogéneo de m ecuaciones lineales con n incógnitas forma un subespacio de F^n. Podemos entonces aplicar la teoría de los espacios vectoriales a este conjunto de soluciones. Por ejemplo, se puede encontrar una base para el espacio solución y cualquier solución puede expresarse como una combinación lineal de los vectores de la base.

Definición. *Se dice que un sistema* $AX = B$ *de* m *ecuaciones con* n *incógnitas es* homogéneo *si* $B = 0$; *de lo contrario se dice que el sistema es* no homogéneo.

Todo sistema homogéneo tiene al menos una solución, a saber,

$$s = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix}.$$

Esta solución se denomina la *solución trivial*. El siguiente resultado proporciona más información sobre el conjunto de soluciones de un sistema homogéneo.

Teorema 3.7. *Sea* $AX = 0$ *un sistema homogéneo de ecuaciones lineales con* n *incógnitas sobre un campo* F. *Sea* K *el conjunto de todas las soluciones para* $AX = 0$. *Entonces* $K = N(L_A)$; *por lo tanto,* K *es un subespacio de* F^n *de dimensión* $n - \textit{rango}(L_A) = n - \textit{rango}(A)$.

DEMOSTRACIÓN. $K = \{s \in F^n\colon\ As = 0\} = N(L_A)$. La segunda parte se sigue del Teorema 2.3. ■

Corolario. *Si* $m < n$ *el sistema* $AX = 0$ *tiene una solución no trivial.*

DEMOSTRACIÓN. Supóngase que $m < n$. Entonces $\text{rango}(A) = \text{rango}(L_A) \leq m$, y entonces $\dim(K) = n - \text{rango}(L_A) \geq n - m > 0$. Como $\dim(K) > 0$, $K \neq \{0\}$. Luego, existe $s \in K$, $s \neq 0$. Entonces s es una solución no trivial para $AX = 0$. ■

Ejemplo 10.

(a) Considérese el sistema

$$\begin{cases} x_1 + 2x_2 + x_3 = 0 \\ x_1 - x_2 - x_3 = 0. \end{cases}$$

Sea

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & -1 & -1 \end{pmatrix}$$

la matriz de coeficientes. Es evidente que $\text{rango}(A) = 2$. Si K es el conjunto solución del sistema, entonces $\dim(K) = 3 - 2 = 1$. Luego, cualquier solución no nula será una base para K. Por ejemplo, como

$$\begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}$$

es una solución,

$$\left\{ \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} \right\}$$

es una base. Entonces cualquier elemento de K es de la forma

$$t\begin{pmatrix}1\\-2\\3\end{pmatrix}=\begin{pmatrix}t\\-2t\\3t\end{pmatrix},$$

donde $t \in R$.

(b) Considérese al sistema $x_1 - 2x_2 + x_3 = 0$ de una ecuación con tres incógnitas. Si $A = (1, -2, 1)$ es la matriz de coeficientes, rango$(A) = 1$. Por lo tanto, si K es el conjunto solución, $\dim(K) = 3 - 1 = 2$. Nótese que

$$\begin{pmatrix}2\\1\\0\end{pmatrix} \quad \text{y} \quad \begin{pmatrix}1\\0\\-1\end{pmatrix}$$

son elementos de K linealmente independientes. Por lo tanto constituyen una base para K, tal que

$$K=\left\{t_1\begin{pmatrix}2\\1\\0\end{pmatrix}+t_2\begin{pmatrix}1\\0\\-1\end{pmatrix}:\quad t_1, t_2 \in R\right\}.$$

En la Sección 3.4 expondremos métodos explícitos de cálculo para encontrar una base para el conjunto solución de un sistema homogéneo.

Ahora pasaremos al estudio de los sistemas no homogéneos. Nuestro siguiente resultado muestra que el conjunto de soluciones de un sistema no homogéneo $AX = B$ puede expresarse en términos del conjunto de soluciones del sistema homogéneo $AX = 0$. Nos referiremos a la ecuación $AX = 0$ como al *sistema homogéneo correspondiente a* AX = B.

Teorema 3.8. *Sea* K *el conjunto solución de un sistema de ecuaciones lineales* AX = B *y sea* K_H *el conjunto solución del sistema homogéneo correspondiente* AX = 0. *Entonces para cualquier solución* s *de* AX = B

$$K = \{s\} + K_H = \{s + k: \quad k \in K_H\}.$$

DEMOSTRACIÓN. Sea s cualquier solución de $AX = B$. Demostraremos que $K = \{s\} + K_H$. Si $w \in K$, entonces $Aw = B$. De aquí que $A(w - s) = Aw - As = B - B = 0$. Entonces $w - s \in K_H$. Luego, existe $k \in K_H$ tal que $w - s = k$, de manera que $w = s + k \in \{s\} + K_H$, y por lo tanto

$$K \subseteq \{s\} + K_H.$$

Recíprocamente, supóngase que $w \in \{s\} + K_H$ entonces $w = s + k$ para alguna $k \in K$. Pero entonces $Aw = A(s + k) = As + Ak = B + 0 = B$, de manera que $w \in K$. Por lo tanto $\{s\} + K_H \subseteq K$ y entonces $K = \{s\} + K_H$. ■

Ejemplo 11.

(a) Considérese el sistema

$$\begin{cases} x_1 + 2x_2 + x_3 = 7 \\ x_1 - x_2 - x_3 = -4. \end{cases}$$

El sistema homogéneo correspondiente al anterior es el sistema dado en el Ejemplo 10(a). Puede comprobarse fácilmente que

$$s = \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix}$$

es una solución del sistema no homogéneo anterior. Así, el conjunto solución del sistema es

$$K = \left\{ \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix} + t \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} : \quad t \in R \right\}$$

por el Teorema 3.8.

(b) Considérese al sistema $x_1 - 2x_2 + x_3 = 4$. El sistema homogéneo correspondiente a este sistema está dado en el Ejemplo 10(b). Como

$$s = \begin{pmatrix} 4 \\ 0 \\ 0 \end{pmatrix}$$

es una solución de este sistema, el conjunto completo de soluciones K se puede escribir como

$$K = \left\{ \begin{pmatrix} 4 \\ 0 \\ 0 \end{pmatrix} + t_1 \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + t_2 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} : \quad t_1, t_2 \in R \right\}.$$

Aun cuando se haya reservado la Sección 3.4 para métodos de cálculo, el teorema siguiente nos proporciona un medio para calcular las soluciones de ciertos sistemas de ecuaciones.

Teorema 3.9. *Sea* $AX = B$ *un sistema de* n *ecuaciones y* n *incógnitas. Si* A *es invertible, entonces el sistema tiene* exactamente una *solución, que será* $A^{-1}B$. *Inversamente, si el sistema tiene únicamente una solución, entonces* A *es invertible.*

DEMOSTRACIÓN. Supóngase que A es invertible. Sustituyendo $A^{-1}B$ en el sistema, tenemos $A(A^{-1}B) = (AA^{-1})B = B$, de manera que $A^{-1}B$ es una solución. Si s es una solución arbitraria, entonces $As = B$ y al multiplicar ambos lados por A^{-1} se tiene que $s = A^{-1}B$. Por tanto, el sistema tiene una y sólo una solución, que es $A^{-1}B$.

Recíprocamente, supóngase que el sistema tiene exactamente una solución s. Sea K_H el conjunto solución para el sistema homogéneo correspondiente $AX = 0$. Por el Teorema 3.8, $\{s\} + K_H$, pero esto sólo puede tenerse cuando $K_H = \{0\}$. Entonces $N(L_A) = \{0\}$ y por lo tanto A es invertible. ■

Ejemplo 12. Considérese el siguiente sistema de 3 ecuaciones con tres incógnitas:

$$(S) \quad \begin{cases} \phantom{2x_1 + {}} 2x_2 + 4x_3 = 2 \\ 2x_1 + 4x_2 + 2x_3 = 3 \\ 3x_1 + 3x_2 + x_3 = 1. \end{cases}$$

En el Ejemplo 7 calculamos la inversa de la matriz de coeficientes A de este sistema, por lo que (S) tiene exactamente una solución:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = A^{-1}B = \begin{pmatrix} \frac{1}{8} & -\frac{5}{8} & \frac{3}{4} \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{2} \\ \frac{3}{8} & -\frac{3}{8} & \frac{1}{4} \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} -\frac{7}{8} \\ \frac{5}{4} \\ -\frac{1}{8} \end{pmatrix}.$$

Utilizaremos esta técnica para resolver sistemas de ecuaciones lineales cuyas matrices de coeficientes son invertibles en la aplicación que da fin a esta sección.

En el Ejemplo 9(c) vimos un sistema de ecuaciones lineales que no tenía solución. Estableceremos ahora un criterio para determinar cuándo un sistema tiene soluciones. Este criterio involucra el rango de la matriz de coeficientes del sistema $AX = B$ y el rango de la matriz $(A \mid B)$. A la matriz $(A \mid B)$ se le denomina *matriz aumentada del sistema* $AX = B$.

Teorema 3.10. *Sea* $AX = B$ *un sistema de ecuaciones lineales. Entonces el sistema tiene al menos una solución si y sólo si rango*(A) = *rango*$(A \mid B)$.

DEMOSTRACIÓN. Decir que $AX = B$ tiene solución, equivale a decir que $B \in R(L_A)$. En la demostración del Teorema 3.4 vimos que $R(L_A) = L\{A^1, A^2, \ldots, A^n\}$, es el subespacio generado por las columnas de A. Entonces $AX = B$ tiene una solución si y sólo si B pertenece a dicho subespacio. Pero $B \in L\{A^1, A^2, \ldots, A^n\}$ si y sólo si $L\{A^1, A^2, \ldots, A^n\} = L\{A^1, A^2, \ldots, A^n, B\}$. Esta última proposición es equivalente a

$$\dim(L\{A^1, A^2, \ldots, A^n\} = \dim(L\{A^1, A^2, \ldots, A^n, B\}).$$

Y de acuerdo con el Teorema 3.4, la ecuación anterior se reduce a

$$\text{rango}(A) = \text{rango}(A \mid B). \quad ■$$

Ejemplo 13. Recuérdese el sistema de ecuaciones

$$\begin{cases} x_1 + x_2 = 0 \\ x_1 + x_2 = 1 \end{cases}$$

del Ejemplo 9(c).

Puesto que

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{y} \quad (A \mid B) = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix},$$

rango$(A) = 1$ y rango$(A \mid B) = 2$. Como los rangos no son iguales, el sistema no tiene soluciones.

Ejemplo 14. Utilizaremos el Teorema 3.10 para determinar si (3, 3, 2) está en el rango de la transformación lineal T: $\mathsf{R}^3 \to \mathsf{R}^3$ definida por

$$\mathsf{T}(a_1, a_2, a_3) = (a_1 + a_2 + a_3,\ a_1 - a_2 + a_3,\ a_1 + a_3).$$

Pero tenemos que $(3, 3, 2) \in \mathsf{R}(\mathsf{T})$ si y sólo si existe un vector $s = (x_1, x_2, x_3) \in \mathsf{R}^3$ tal que $\mathsf{T}(s) = (3, 3, 2)$. Tal vector s deberá ser una solución del sistema

$$\begin{cases} x_1 + x_2 + x_3 = 3 \\ x_1 - x_2 + x_3 = 3 \\ x_1 \qquad\ \ + x_3 = 2. \end{cases}$$

Como los rangos de la matriz de coeficientes y de la matriz aumentada de este sistema son 2 y 3, respectivamente, se tiene que este sistema no tiene soluciones. Por lo tanto, $(3, 3, 2) \notin \mathsf{R}(\mathsf{T})$.

Una aplicación

En 1973 Wassily Leontief ganó el Premio Nobel de Economía por su trabajo en el desarrollo de un modelo matemático que se puede utilizar para describir diversos fenómenos económicos. Terminaremos esta sección aplicando algunas de las ideas que hemos estudiado para ilustrar dos casos especiales de su trabajo.

Principiaremos considerando una sociedad sencilla constituida por tres personas (empresas): un campesino que produce todos los alimentos, un sastre que hace todo el vestido y un carpintero que construye todo lo de la vivienda. Supondremos que las tres personas venden y compran en un abasto central y que todo lo que se produce se consume. Como ningún producto entra o sale del sistema, en este caso se trata del *modelo cerrado.*

Cada uno de los tres individuos consumirá de cada uno de los tres productos producidos dentro de la sociedad. Supóngase que la proporción de cada uno de los productos consumidos por cada una de las personas está dada en el cuadro siguiente. Nótese que cada una de las columnas del cuadro deben sumar 1.

	Alimentación	*Vestido*	*Vivienda*
Campesino	0.40	0.20	0.20
Sastre	0.10	0.70	0.20
Carpintero	0.50	0.10	0.60

Sean p_1, p_2 y p_3 respectivamente los ingresos del campesino, del sastre y del carpintero. Para tener la certeza de que esta sociedad sobrevive, se requiere que el consumo de cada individuo sea igual a su ingreso. En el caso del campesino, este requisito puede traducirse en la ecuación $0.40p_1 + 0.20p_2 + 0.20p_3 = p_1$. Entonces necesitamos considerar el sistema de ecuaciones lineales

$$\begin{cases} 0.40p_1 + 0.20p_2 + 0.20p_3 = p_1 \\ 0.10p_1 + 0.70p_2 + 0.20p_3 = p_2 \\ 0.50p_1 + 0.10p_2 + 0.60p_3 = p_3 \end{cases}$$

o su equivalente, $AP = P$, donde

$$P = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$$

y A es la matriz de coeficientes del sistema. Dentro de este contexto A se denomina *matriz de consumo* y $AP = P$ la *condición de equilibrio.*

Para matrices B y C del mismo tamaño utilizaremos la notación $B \geq C[B > C]$ para indicar $B_{ij} \geq C_{ij}[B_{ij} > C_{ij}]$ para toda i y j. B se denominará *no negativa* [*positiva*] si $B \geq O[B > O]$, donde O es la matriz nula.

Al principio puede parecer razonable reemplazar la condición de equilibrio por la desigualdad $AP \leq P$, esto es, la condición que el consumo no exceda a la producción. Pero de hecho $AP \leq P$ implica que $AP = P$ en el modelo cerrado, pues de otra manera existiría una k para la cual

$$p_k > \sum_j A_{kj}p_j.$$

Por tanto, como las columnas de A suman 1,

$$\sum_i p_i > \sum_i \sum_j A_{ij}p_j = \sum_j (\sum_i A_{ij})p_j = \sum_j p_j,$$

lo cual es una contradicción.

Una solución del sistema homogéneo $(I - A)X = O$ equivalente a la condición de equilibrio es

$$P = \begin{pmatrix} 0.25 \\ 0.35 \\ 0.40 \end{pmatrix}.$$

Podemos interpretar este hecho como indicando que la sociedad sobrevivirá si el campesino, el sastre y el carpintero tienen ingresos en la proporción 25:35:40 (o 5:7:8).

Nótese que no estamos interesados simplemente en una solución no trivial al sistema, sino en una que sea no negativa. Por ello debemos considerar si el sistema $(I - A)X = O$ tiene o no una solución no negativa,

donde A sea una matriz no negativa cuyas columnas suman 1. Un teorema útil en este aspecto [cuya demostración puede encontrarse en la publicación "Aplicaciones de las Matrices a Modelos Económicos y a las Interrelaciones de las Ciencias Sociales" de Ben Noble, *Proceedings of the Summer Conference for College Teachers on Applied Mathematics* (1971), CUPM, Berkeley, California] se enuncia a continuación.

Teorema 3.11. *Sea* A *una matriz de consumo de* $n \times n$ *de la forma*

$$A = \begin{pmatrix} B & C \\ D & E \end{pmatrix},$$

donde D *es una matriz positiva de* $1 \times (n-1)$ *y* C *es una matriz positiva de* $(n-1) \times 1$. *Entonces* $(I - A)X = 0$ *tiene un conjunto de soluciones unidimensional generado por un vector no negativo.*

Obsérvese que toda matriz de consumo positiva satisface la hipótesis de este teorema. La matriz siguiente también la satisface.

$$\begin{pmatrix} 0.75 & 0.50 & 0.65 \\ 0 & 0.25 & 0.35 \\ 0.25 & 0.25 & 0 \end{pmatrix}$$

En el *modelo abierto* suponemos que existe una demanda externa para cada uno de los productos producidos. Volviendo a nuestra sociedad sencilla, sean x_1, x_2 y x_3 las cantidades de alimento, vestido y vivienda producidas en función de las demandas externas d_1, d_2 y d_3. Sea A la matriz de 3×3 tal que A_{ij} representa la proporción del producto i consumido en producir el producto j. Entonces el superávit de alimentos en la sociedad es

$$x_1 - (A_{11}x_1 + A_{12}x_2 + A_{13}x_3),$$

esto es, la cantidad de alimentos producidos menos la cantidad de alimentos consumidos para producir los tres productos. La suposición de que todo lo producido se consume nos da una condición de equilibrio similar para el modelo abierto, esto es, que los superávits de cada uno de los tres productos deben ser iguales a las correspondientes demandas externas. Por lo tanto

$$x_i - \sum_{j=1}^{3} A_{ij}x_j = d_i \qquad \text{para } j = 1, 2 \text{ y } 3.$$

En general, debemos encontrar una solución no negativa para el sistema $(I - A)X = D$, donde A y D son matrices no negativas y la suma de los elementos de las columnas de A no es mayor que uno. Es fácil ver que si $(I - A)^{-1}$ existe y es no negativa, entonces la solución deseada será $(I - A)^{-1}D$.

Recuérdese que para un número real a, la serie $1 + a + a^2 + \ldots$ converge a $(1 - a)^{-1}$ si $|a| < 1$. De la misma manera puede demostrarse

(utilizando el concepto de convergencia de matrices desarrollado en la Sección 5.3) que la serie $I + A + A^2 + \ldots$ converge a $(I - A)^{-1}$ si A^n converge a la matriz nula. En este caso $(I - A)^{-1}$ será no negativa puesto que las matrices $I, A, A^2, \ldots$ son no negativas.

Para ilustrar el modelo abierto, supóngase que el 30% de los alimentos se utiliza para producir alimentos, 20% para producir vestido y 30% para la vivienda. De la misma forma, supóngase que el 10% del vestido se destina a la producción de alimentos, 40% para producir vestido y 10% para la vivienda. Finalmente, supóngase que 30% de la vivienda se utiliza para producir alimentos, 20% para producir vestido y 30% para la producción de la vivienda. Entonces, la matriz de consumo es

$$A = \begin{pmatrix} 0.30 & 0.20 & 0.30 \\ 0.10 & 0.40 & 0.10 \\ 0.30 & 0.20 & 0.30 \end{pmatrix},$$

y entonces

$$I - A = \begin{pmatrix} 0.70 & -0.20 & -0.30 \\ -0.10 & 0.60 & -0.10 \\ -0.30 & -0.20 & 0.70 \end{pmatrix} \quad \text{y} \quad (I - A)^{-1} = \begin{pmatrix} 2.0 & 1.0 & 1.0 \\ 0.5 & 2.0 & 0.5 \\ 1.0 & 1.0 & 2.0 \end{pmatrix}.$$

Como $(I - A)^{-1}$ es no negativa podemos encontrar una solución no negativa (única) para $(I - A)X = D$ para cualquier demanda D. Por ejemplo si

$$D = \begin{pmatrix} 30 \\ 20 \\ 10 \end{pmatrix},$$

entonces

$$X = (I - A)^{-1}D = \begin{pmatrix} 90 \\ 60 \\ 70 \end{pmatrix}.$$

Entonces, se debe producir un volumen de producción de 90 unidades de alimentos, 60 unidades de vestido y 70 unidades de vivienda para satisfacer una demanda de 30 unidades de alimentos, 20 unidades de vestido y 10 unidades de vivienda.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

   (a) Cualquier sistema de ecuaciones lineales tiene al menos una solución.
   (b) Todo sistema de ecuaciones lineales tiene como máximo una solución.

(c) Cualquier sistema homogéneo de ecuaciones lineales tiene al menos una solución.
(d) Cualquier sistema de n ecuaciones lineales con n incógnitas tiene como máximo una solución.
(e) Cualquier sistema de n ecuaciones lineales con n incógnitas tiene al menos una solución.
(f) Si el sistema homogéneo correspondiente a un sistema de ecuaciones lineales dado tiene una solución, entonces el sistema dado tiene una solución.
(g) Si la matriz de coeficientes de un sistema homogéneo de n ecuaciones lineales con n incógnitas es invertible, entonces el sistema carece de soluciones triviales.
(h) El conjunto solución de cualquier sistema de m ecuaciones lineales con n incógnitas es un subespacio de F^n.

2. Para cada uno de los siguientes sistemas de ecuaciones lineales, encontrar la dimensión y una base para el conjunto solución.

(a) $\begin{cases} x_1 + x_2 - x_3 = 0 \\ 4x_1 + x_2 - 2x_3 = 0 \end{cases}$

(b) $\begin{cases} 2x_1 + x_2 - x_3 = 0 \\ x_1 - x_2 + x_3 = 0 \\ x_1 + 2x_2 - 2x_3 = 0 \end{cases}$

(c) $x_1 + 2x_2 - 3x_3 + x_4 = 0$

(d) $\begin{cases} x_1 + 2x_2 = 0 \\ x_1 - x_2 = 0 \end{cases}$

3. Utilizando los resultados del Ejercicio 2 encontrar todas las soluciones de los siguientes sistemas.

(a) $\begin{cases} x_1 + x_2 - x_3 = 1 \\ 4x_1 + x_2 - 2x_3 = 3 \end{cases}$

(b) $\begin{cases} 2x_1 + x_2 - x_3 = 5 \\ x_1 - x_2 + x_3 = 1 \\ x_1 + 2x_2 - 2x_3 = 4 \end{cases}$

(c) $x_1 + 2x_2 - 3x_3 + x_4 = 1$

(d) $\begin{cases} x_1 + 2x_2 = 5 \\ x_1 - x_2 = -1 \end{cases}$

4. Sea A la matriz de coeficientes de

$$\begin{cases} x_1 + 2x_2 - x_3 = 5 \\ x_1 + x_2 + x_3 = 1 \\ 2x_1 - 2x_2 + x_3 = 4. \end{cases}$$

(a) Demostrar que A es invertible.
(b) Calcular A^{-1}.
(c) Utilizar A^{-1} para resolver el sistema.

5. Dar un ejemplo de un sistema de n ecuaciones con n incógnitas con un número infinito de soluciones.

6. Sea $\mathsf{T}\colon \mathsf{R}^3 \to \mathsf{R}^2$ definida por $\mathsf{T}(a, b, c) = (a + b, 2a - c)$. Describir $\mathsf{T}^{-1}\{(1, 11)\}$.

7. Determinar cuál de los siguientes sistemas de ecuaciones lineales tiene solución.

(a) $$\begin{cases} x_1 + x_2 - x_3 + 2x_4 = 2 \\ x_1 + x_2 + 2x_3 = 1 \\ 2x_1 + 2x_2 + x_3 + 2x_4 = 4 \end{cases}$$

(b) $$\begin{cases} x_1 + x_2 - x_3 = 1 \\ 2x_1 + x_2 + 3x_3 = 2 \end{cases}$$

(c) $$\begin{cases} x_1 + 2x_2 + 3x_3 = 1 \\ x_1 + x_2 - x_3 = 0 \\ x_1 + 2x_2 + x_3 = 3 \end{cases}$$

(d) $$\begin{cases} x_1 + x_2 + 3x_3 - x_4 = 0 \\ x_1 + x_2 + x_3 + x_4 = 1 \\ x_1 - 2x_2 + x_3 - x_4 = 1 \\ 4x_1 + x_2 + 8x_3 - x_4 = 0 \end{cases}$$

8. Demostrar que un sistema $AX = B$ de m ecuaciones lineales con n incógnitas tiene solución si y sólo si $B \in \mathsf{R}(\mathsf{L}_A)$.

9. Demostrar o dar un contraejemplo al siguiente enunciado: Si la matriz de coeficientes de un sistema de m ecuaciones lineales con n incógnitas tiene rango m, entonces el sistema tiene una solución.

10. En el modelo cerrado de Leontief con alimentación, vestido y vivienda como industrias básicas, supóngase que la matriz de consumo es

$$A = \begin{pmatrix} \frac{7}{16} & \frac{1}{2} & \frac{3}{16} \\ \frac{5}{16} & \frac{1}{6} & \frac{5}{16} \\ \frac{1}{4} & \frac{1}{3} & \frac{1}{2} \end{pmatrix}.$$

¿En qué proporción deben producir el campesino, el sastre y el carpintero de manera que se alcance el equilibrio?

11. En la representación del modelo abierto de Leontief, supóngase que

$$A = \begin{pmatrix} \frac{1}{2} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{5} \end{pmatrix}$$

y que el vector de demanda es $D = \begin{pmatrix} 2 \\ 5 \end{pmatrix}$. ¿Cuánto de cada producto se deberá producir para satisfacer esta demanda?

3.4 SISTEMAS DE ECUACIONES LINEALES: ASPECTOS DE CALCULO

En la Sección 3.3 obtuvimos una condición necesaria y suficiente para que un sistema de ecuaciones lineales tenga soluciones (Teorema 3.10) y

aprendimos cómo expresar las soluciones de un sistema no homogéneo en términos de las soluciones del sistema homogéneo correspondiente (Teorema 3.8). Este último resultado nos permite determinar todas las soluciones de un sistema dado si podemos encontrar una solución de dicho sistema y una base para el conjunto solución del sistema homogéneo correspondiente. En esta sección utilizaremos las operaciones elementales con renglones para alcanzar estos dos objetivos. La esencia de esta técnica es transformar un sistema dado de ecuaciones lineales en un sistema que tenga las mismas soluciones pero que sea más fácil de resolver (como en la Sección 1.4).

Definición. *Dos sistemas de* m *ecuaciones lineales con* n *incógnitas se llaman* equivalentes *si tienen el mismo conjunto de soluciones.*

El siguiente teorema y su corolario nos proporcionan un método útil para obtener sistemas equivalentes.

Teorema 3.12. *Sea* (S): $AX = B$ *un sistema de* m *ecuaciones lineales con* n *incógnitas y sea* C *cualquier matriz invertible de* $m \times m$. *Entonces el sistema* (S′): $(CA)X = CB$ *es equivalente a* (S).

DEMOSTRACIÓN. Sea K el conjunto de soluciones para (S) y K' el conjunto de soluciones para (S'). Si $w \in K$, entonces $Aw = B$. Luego entonces, $CAw = CB$ y por lo tanto $w \in K'$. Entonces $K \subseteq K'$.

Recíprocamente, si $w \in K'$, entonces $CAw = CB$. En consecuencia, $Aw = C^{-1}(CAw) = C^{-1}(CB) = B$ y así $w \in K$. Entonces $K' \subseteq K$, y por lo tanto $K = K'$. ■

Corolario. *Sea* $AX = B$ *un sistema de* m *ecuaciones lineales con* n *incógnitas. Si* $(A' \mid B')$ *se obtiene a partir de* $(A \mid B)$ *mediante un número finito de operaciones elementales con renglones, entonces el sistema* $A'X = B'$ *es equivalente al sistema original.*

DEMOSTRACIÓN. Supóngase que $(A' \mid B')$ se obtiene de $(A \mid B)$ por medio de operaciones elementales con renglones. Estas se pueden realizar multiplicando por matrices elementales de $m \times m$ $E_1, \ldots, E_p$. Sea $C = E_p \ldots E_1$; entonces $(A' \mid B') = C(A \mid B) = (CA \mid CB)$, y como todas las E_i son invertibles también lo es C. Ahora bien, $A' = CA$ y $B' = CB$. Luego, por el Teorema 3.12, el sistema $A'X = B'$ es equivalente al sistema $AX = B$. ■

Ejemplo 15. Para encontrar todas las soluciones de

$$\begin{cases} x_1 + 2x_2 + x_3 - x_4 = 2 \\ x_1 + x_2 + x_3 = 3 \\ 3x_1 + 2x_2 + 3x_3 - 2x_4 = 1, \end{cases}$$

construimos la matriz aumentada

$$\begin{pmatrix} 1 & 2 & 1 & -1 & 2 \\ 1 & 1 & 1 & 0 & 3 \\ 3 & 2 & 3 & -2 & 1 \end{pmatrix}$$

del sistema y la simplificamos mediante una secuencia de operaciones elementales con renglones de la siguiente manera.

(a) Colocamos 1 en el primer renglón, primera columna. (Esto es ya el caso.)

(b) Por medio de operaciones del tipo 3, utilizamos el primer renglón para obtener ceros en las posiciones restantes de la primera columna. La matriz que se obtiene es

$$\begin{pmatrix} 1 & 2 & 1 & -1 & 2 \\ 0 & -1 & 0 & 1 & 1 \\ 0 & -4 & 0 & 1 & -5 \end{pmatrix}.$$

Para las operaciones restantes ya no se utiliza el primer renglón.

(c) Luego (utilizando los renglones restantes) colocamos un 1 en el segundo renglón y en la columna lo más a la izquierda posible —en este caso la segunda columna. Entonces hacemos operaciones con renglones del tipo 3 para obtener ceros en las posiciones restantes de esta columna. Estas operaciones dan

$$\begin{pmatrix} 1 & 0 & 1 & 1 & 4 \\ 0 & 1 & 0 & -1 & -1 \\ 0 & 0 & 0 & -3 & -9 \end{pmatrix}.$$

(d) Para terminar, utilizando únicamente el renglón restante, colocamos un 1 en el tercer renglón y la columna lo más a la izquierda posible, en este caso la cuarta columna. Por medio de operaciones del tipo 3, usamos este 1 para producir ceros en la cuarta columna, y así obtenemos

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 3 \end{pmatrix}.$$

Esta última matriz puede ser traducida a un sistema de ecuaciones

$$\begin{cases} x_1 + \quad x_3 \quad = 1 \\ \qquad x_2 \qquad = 2 \\ \qquad\qquad x_4 = 3 \end{cases}$$

equivalente al sistema dado. Evidentemente $x_2 = 2$ y $x_4 = 3$, pero x_1 y x_3 pueden tener cualquier valor siempre que su suma sea 1. Haciendo

$x_3 = t$, tenemos entonces que $x_1 = 1 - t$. Luego, una solución arbitraria tiene la forma

$$\begin{pmatrix} 1-t \\ 2 \\ t \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 3 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}.$$

Obsérvese que

$$\left\{\begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix}\right\}$$

es una base para el sistema de ecuaciones homogéneo correspondiente al sistema dado.

En el ejemplo anterior realizamos operaciones elementales con renglones en la matriz aumentada del sistema hasta obtener la matriz aumentada de un sistema con las propiedades 1, 2 y 3 dadas en la página 29. Esa matriz tiene un nombre especial.

Definición. *Se dice que una matriz es* escalonada por renglones *si se satisfacen las tres condiciones siguientes:*

(a) *Cualquier renglón que tenga un elemento no nulo precederá a cualquier renglón (si es que existe alguno) donde todos los elementos sean ceros.*

(b) *El primer elemento no nulo de cada renglón es el único elemento no nulo en su columna.*

(c) *El primer elemento no nulo en cada renglón es* 1 *y aparece en una columna que está a la derecha del* 1 *que encabeza a cualquier renglón anterior.*

Ejemplo 16.

(a) La primera matriz del Ejemplo 15(d) es una matriz escalonada por renglones. Nótese que el primer elemento no nulo de cada renglón es 1 y que la columna que lo contiene tiene ceros en el resto de los elementos. También véase que cada vez que nos desplazamos hacia abajo a un nuevo renglón, debemos desplazarnos cuando menos una (y posiblemente más) columna(s) hacia la derecha para encontrar el primer elemento no nulo del nuevo renglón.

(b) Las siguientes matrices *no* son escalonadas por renglones:

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix},$$

porque la primera columna contiene más de un elemento no nulo;

$$\begin{pmatrix} 0 & 1 & 0 & 2 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix},$$

porque el primer elemento no nulo del segundo renglón no está a la derecha del primer elemento no nulo del primer renglón; y

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

porque el primer elemento no nulo del primer renglón no es 1.

La facilidad con que resolvimos el sistema final de ecuaciones del Ejemplo 15, se debe al hecho de que la matriz aumentada de este sistema es una matriz escalonada por renglones. Presentaremos en seguida un procedimiento para resolver cualquier sistema de ecuaciones lineales para el cual la matriz aumentada sea una matriz escalonada por renglones. Sin embargo, primero estableceremos que toda matriz puede ser transformada en una matriz escalonada por renglones mediante operaciones elementales con renglones.

Teorema 3.13. *Toda matriz puede ser transformada en una matriz escalonada por renglones por medio de un número finito de operaciones elementales con renglones.*

DEMOSTRACIÓN. La demostración se hará por inducción sobre el número de columnas de la matriz. Dejaremos como ejercicio la demostración del resultado para matrices de una columna. Supóngase que la conclusión es válida para matrices que tienen n columnas, para algún entero $n \geq 1$, y sea A una matriz de $m \times (n + 1)$. Escríbase A en la forma $A = (A' \mid B)$, donde B es la última columna de A y A' es la matriz de $m \times n$ obtenida al suprimir la última columna de A. Por la hipótesis de inducción, A' puede ser transformada en una matriz Q escalonada por renglones por medio de un número finito de operaciones elementales con renglones. Sea C el producto de las matrices elementales que corresponden a estas operaciones con renglones. Entonces

$$CA = C(A' \mid B) = (CA' \mid CB) = (Q \mid B'),$$

donde $B' = CB$. Claramente $(Q \mid B')$ es una matriz escalonada por renglones a menos que contenga un renglón de la forma $(0 \ldots 0\ a)$ donde $a \neq 0$. Multiplicando a tal renglón por a^{-1}, sumando múltiplos adecuados de este renglón a los otros renglones y realizando el intercambio adecuado de renglones, para alguna j podemos transformar $(Q \mid B')$ en una matriz $(Q \mid e_j)$ escalonada por renglones por medio de un número finito de operaciones elementales con renglones. Esto completa la inducción. ■

Puede demostrarse (véase el Ejercicio 9) que a toda matriz le corresponde una *única* matriz escalonada por renglones; esto es, si por distintas secuencias de operaciones elementales con renglones se transforma a la matriz en matrices Q y Q', ambas escalonadas por renglones, entonces $Q = Q'$.

Describiremos ahora un método para resolver un sistema en el que la matriz aumentada sea una matriz escalonada por renglones. Para ilustrar el procedimiento, consideremos al sistema

$$\begin{cases} 2x_1 + 3x_2 + x_3 + 4x_4 - 9x_5 = 17 \\ x_1 + x_2 + x_3 + x_4 - 3x_5 = 6 \\ x_1 + x_2 + x_3 + 2x_4 - 5x_5 = 8 \\ 2x_1 + 2x_2 + 2x_3 + 3x_4 - 8x_5 = 14 \end{cases}$$

para el que la matriz aumentada es

$$\begin{pmatrix} 2 & 3 & 1 & 4 & -9 & 17 \\ 1 & 1 & 1 & 1 & -3 & 6 \\ 1 & 1 & 1 & 2 & -5 & 8 \\ 2 & 2 & 2 & 3 & -8 & 14 \end{pmatrix}.$$

La siguiente secuencia de matrices obtenida por operaciones con renglones ilustra cómo se transforma a la matriz aumentada en una matriz escalonada por renglones:

$$\begin{pmatrix} 1 & 1 & 1 & 1 & -3 & 6 \\ 2 & 3 & 1 & 4 & -9 & 17 \\ 1 & 1 & 1 & 2 & -5 & 8 \\ 2 & 2 & 2 & 3 & -8 & 14 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 & 1 & -3 & 6 \\ 0 & 1 & -1 & 2 & -3 & 5 \\ 0 & 0 & 0 & 1 & -2 & 2 \\ 0 & 0 & 0 & 1 & -2 & 2 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 0 & 2 & -1 & 0 & 1 \\ 0 & 1 & -1 & 2 & -3 & 5 \\ 0 & 0 & 0 & 1 & -2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 2 & 0 & -2 & 3 \\ 0 & 1 & -1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & -2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

El sistema de ecuaciones (equivalente al original) asociado con esta última matriz (considerada como matriz aumentada) es

$$\begin{cases} x_1 \quad + 2x_3 \quad - 2x_5 = 3 \\ \quad x_2 - \ x_3 \quad + \ x_5 = 1 \\ \quad\quad\quad x_4 - 2x_5 = 2. \end{cases}$$

Nótese que hemos ignorado al último renglón pues está totalmente formado por ceros.

Para resolver un sistema para el cual la matriz aumentada toma la forma escalonada, divídase a las variables $x_1, x_2, \ldots, x_5$ en dos conjuntos. El primer conjunto consta de cada una de las variables que aparecen más a la izquierda en las ecuaciones del sistema (en este caso el conjunto es $\{x_1, x_2, x_4\}$). El segundo conjunto consta del resto de las variables (en este caso, $\{x_3, x_5\}$). A cada variable del segundo conjunto se le asigna un valor paramétrico $t_1, t_2, \ldots$ ($x_3 = t_1$, $x_5 = t_2$) y se resuelve para las variables del primer conjunto en términos de las del segundo:

$$\begin{aligned} x_1 &= -2x_3 + 2x_5 + 3 = -2t_1 + 2t_2 + 3 \\ x_2 &= \quad x_3 - \ x_5 + 1 = \quad t_1 - \ t_2 + 1 \\ x_4 &= \quad\quad 2x_5 + 2 = \quad\quad 2t_2 + 2. \end{aligned}$$

Así, una solución arbitraria, s, es de la forma

$$s = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -2t_1 + 2t_2 + 3 \\ t_1 - t_2 + 1 \\ t_1 \\ 2t_2 + 2 \\ t_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} + t_1 \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t_2 \begin{pmatrix} 2 \\ -1 \\ 0 \\ 2 \\ 1 \end{pmatrix},$$

donde $t_1, t_2 \in R$. Nótese que

$$\left\{ \begin{pmatrix} -2 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 0 \\ 2 \\ 1 \end{pmatrix} \right\}$$

forma una base para el conjunto de soluciones correspondiente al sistema de ecuaciones homogéneo.

Para utilizar este procedimiento para resolver un sistema de m ecuaciones con n incógnitas $AX = B$, véase primero si rango(A) = rango $(A \mid B)$. Si esta igualdad no se satisface, entonces el sistema no tiene soluciones. En seguida (siempre que el sistema tenga soluciones) utilícense operaciones elementales con renglones para transformar la matriz aumen-

tada $(A \mid B)$ en una matriz escalonada por renglones $(A' \mid B')$. Descártense los renglones nulos en $(A' \mid B')$ y escríbase de nuevo el sistema de ecuaciones asociado con $(A' \mid B')$. Resuélvase el sistema como se hizo anteriormente y se obtendrá una solución arbitraria de la forma

$$s = s_0 + t_1u_1 + t_2u_2 + \ldots + t_{n-m'}u_{n-m'},$$

donde m' es el número de renglones no nulos en A', $(m' \leq m)$. La ecuación anterior sugiere que una solución arbitraria, s, puede expresarse en términos de $n - m'$ parámetros. El teorema siguiente establece que s no puede expresarse en menos de $n - m'$ parámetros.

Teorema 3.14. *Sea* $AX = B$ *un sistema de* m *ecuaciones no nulas con* n *incógnitas. Supóngase que rango*$(A) =$ *rango*$(A \mid B)$ *y que* $(A \mid B)$ *es una matriz escalonada por renglones. Entonces*

(a) *rango*$(A) = m$.

(b) *Si la solución general, obtenida por el procedimiento anterior es de la forma*

$$s = s_0 + t_1u_1 + t_2u_2 + \ldots + t_{n-m}u_{n-m},$$

entonces $\{u_1, u_2, \ldots, u_{n-m}\}$ *es una base para el conjunto de soluciones del sistema homogéneo correspondiente y* s_0 *es una solución del sistema original.*

DEMOSTRACIÓN. Como $(A \mid B)$ es una matriz escalonada por renglones, rango$(A \mid B) =$ rango$(A) = m$, de acuerdo con los Ejercicios 5 y 6.

Sea K el conjunto de soluciones para $AX = B$ y K_H el conjunto de soluciones para $AX = 0$. Haciendo $t_1 = t_2 = \ldots = t_{n-m} = 0$, $s = s_0 \in K$. Pero por el Teorema 3.8, $K = \{s_0\} + K_H$, por lo que $K_H = K - \{s_0\} = L(\{u_1, u_2, \ldots, u_{n-m}\})$.

Como rango$(A) = m$, $\dim(K_H) = n - m$. Entonces como $\dim(K_H) = n - m$ y K_H es generado por un conjunto $\{u_1, u_2, \ldots, u_{n-m}\}$ que contiene a lo más $n - m$ elementos, concluimos que el conjunto anterior es una base para K_H. ∎

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

(a) Si $(A' \mid B')$ se obtiene a partir de $(A \mid B)$ mediante una secuencia finita de operaciones elementales con columnas, entonces los sistemas $AX = B$ y $A'X = B'$ son equivalentes.

(b) Si $(A' \mid B')$ se obtiene a partir de $(A \mid B)$ mediante una secuencia finita de operaciones elementales con renglones, entonces los sistemas $AX = B$ y $A'X = B'$ son equivalentes.

(c) Si A es una matriz de $n \times n$ de rango n, entonces la matriz escalonada por renglones de A es I_n.
(d) Cualquier matriz se puede convertir en una matriz escalonada por renglones por medio de una secuencia finita de operaciones elementales con renglones.
(e) Si $(A \mid B)$ es una matriz escalonada por renglones, entonces el sistema $AX = B$ debe de tener una solución.
(f) Sea $AX = B$ un sistema de m ecuaciones lineales con n incógnitas para el cual la matriz aumentada es una matriz escalonada por renglones. Si este sistema tiene solución, entonces la dimensión del conjunto de soluciones de $AX = 0$ es $n - m'$, donde m' es igual al número de renglones no nulos de A.
(g) Si una matriz A se transforma por medio de operaciones elementales con renglones en la matriz escalonada por renglones A', entonces el número de renglones no nulos en A' es igual al rango de A.

2. Encontrar todas las soluciones a los sistemas de ecuaciones en los Ejercicios 2, 3 y 4 de la Sección 3.3 mediante la técnica usada en esta sección.

3. Supóngase que la matriz aumentada del sistema $AX = B$ se transforma en la matriz escalonada $(A' \mid B')$ mediante una secuencia finita de operaciones elementales con renglones.

(a) Demostrar que $\text{rango}(A') \neq \text{rango}((A' \mid B'))$ si y sólo si $(A' \mid B')$ contiene un renglón en donde el único elemento no nulo queda ubicado en la última columna.
(b) Deducir que $AX = B$ tiene soluciones si y sólo si $(A' \mid B')$ no contiene ningún renglón en el cual el único elemento no nulo está ubicado en la última columna.

4. Para cada uno de los siguientes sistemas, aplicar el Ejercicio 3 para determinar si el sistema tiene soluciones. Si existen soluciones, encontrarlas todas. Finalmente, encontrar una base para los sistemas homogéneos correspondientes.

(a) $$\begin{cases} x_1 + 2x_2 - x_3 + x_4 = 2 \\ 2x_1 + x_2 + x_3 - x_4 = 3 \\ x_1 + 2x_2 - 3x_3 + 2x_4 = 2 \end{cases}$$

(b) $$\begin{cases} x_1 + x_2 - 3x_3 + x_4 = -2 \\ x_1 + x_2 + x_3 - x_4 = 2 \\ x_1 + x_2 - x_3 = 0 \end{cases}$$

(c) $$\begin{cases} x_1 + x_2 - 3x_3 + x_4 = 1 \\ x_1 + x_2 + x_3 - x_4 = 2 \\ x_1 + x_2 - x_3 = 0 \end{cases}$$

5. Demostrar que si A es una matriz escalonada por renglones, entonces $\text{rango}(A)$ es igual al número de renglones no nulos de A.

6. Si $(A \mid B)$ es una matriz escalonada por renglones, demostrar que A también es escalonada por renglones.

7. Demostrar el Teorema 3.13 para matrices de una sola columna.

8. Demostrar el Teorema 3.13 de la manera siguiente. Sea A una matriz de $n \times n$ de rango r. Como el resultado es inmediato para $r = 0$, supóngase $r > 0$. Sea β la base ordenada estándar para F^n, defínase $\mathsf{W}_k = L\{\mathsf{L}_A(e_1), \mathsf{L}_A(e_2), \ldots, \mathsf{L}_A(e_k)\}$ para $1 \leq k \leq n$ y defínase $k_j = \min\{i: \dim(\mathsf{W}_i) = j\}$ para $1 \leq j \leq r$. Demostrar que $k_1 < k_2 < \ldots < k_r$ y que $k_j \geq j$ para toda j. Sea $z_j = \mathsf{L}_A(e_{k_j})$, y demostrar que $\{z_1, z_2, \ldots, z_r\}$ es linealmente independiente. Extender este conjunto a una base β' para F^m. Hacer $B = [\mathsf{L}_A]_\beta^{\beta'}$ y demostrar lo siguiente:

   (a) $B = CA$ para alguna matriz invertible C de $m \times m$.
   (b) B es escalonada por renglones.
   (c) B se puede obtener a partir de A mediante un número finito de operaciones elementales con renglones.

9. (a) Demostrar que si Q y Q' son matrices de $m \times n$ escalonadas por renglones tales que Q puede ser transformada en Q' por medio de un número finito de operaciones elementales con renglones, entonces $Q = Q'$. *Sugerencias:* Emplear inducción sobre n.
   (b) Deducir que si A es cualquier matriz, entonces existe una única matriz escalonada por renglones que puede obtenerse a partir de A mediante un número finito de operaciones elementales con renglones.

INDICE DE LAS DEFINICIONES DEL CAPITULO 3

Capítulo 4

Determinantes

Durante un tiempo los determinantes jugaron un papel fundamental en el estudio del álgebra lineal; ahora, sin embargo, tienen una importancia mucho menor. Veremos, de hecho, que virtualmente nuestra única utilización de los determinantes será en el cálculo de los "eigenvalores". Por esta razón las cuestiones más importantes que serán necesarias para los próximos capítulos se resumen en la Sección 4.5, de manera que el lector que no esté interesado en seguir un desarrollo de la teoría de los determinantes podrá pasar inmediatamente a dicha sección.

El determinante de una matriz cuadrada con elementos de un campo F es un escalar (elemento de F), por lo que podemos considerar al determinante como una función cuyo dominio es $\mathsf{M}_{n\times n}(F)$ y que toma valores de F. Aun cuando el determinante de una matriz cuadrada pueda ser definido en términos de los elementos de la matriz, la definición resultante es comprometedora para ser utilizada en operaciones. En vez de definir el determinante de esta manera, en la Sección 4.2 definiremos al determinante como una función δ: $\mathsf{M}_{n\times n}(F) \to F$ que tiene tres propiedades importantes. En dicha sección también verificaremos que el método corriente para evaluar un determinante mediante la expansión a lo largo de una columna es, de hecho, un determinante en el sentido de nuestra definición. La Sección 4.3 contiene otras propiedades adicionales de los determinantes y demuestra que existe un determinante único en $\mathsf{M}_{n\times n}(F)$, es decir, que las tres propiedades que definen un determinante son satisfechas por una y sólo una función de $\mathsf{M}_{n\times n}(F)$ en F. La Sección 4.4 utiliza los determinantes para encontrar la inversa de una matriz invertible y para resolver un sistema de ecuaciones lineales que posee una matriz de coeficientes invertibles por medio de la regla de Cramer.

El capítulo principia con una exposición de la teoría general en una forma sencilla. En esta sección investigaremos también el significado geométrico de los determinantes en términos de área y orientación. Los lectores que hayan estudiado cálculo avanzado recordarán que un cambio de coordenadas en las integrales múltiples requería del uso de un determinante llamado jacobiano.

4.1 DETERMINANTES DE ORDEN 2

Algunas veces asignaremos a toda matriz de $n \times n$ con elementos de un campo F un escalar llamado "el determinante" de la matriz, pero primero consideraremos un caso especial fácil.

Definición. *El* determinante *de una matriz* A *de* 2×2 *con elementos de un campo* F *es el escalar* $A_{11}A_{22} - A_{12}A_{21}$, *que denotaremos por* det(A).

Ejemplo 1. Considérese el siguiente elemento de $M_{2\times 2}(R)$:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

Entonces

$$\det(A) = 1 \cdot 4 - 2 \cdot 3 = -2.$$

En la exposición siguiente será conveniente representar una matriz A de 2×2 en términos de sus renglones; como anteriormente, escribiremos

$$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$$

y representaremos su determinante mediante

$$\det \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}.$$

El determinante tiene las siguientes propiedades importantes.

Teorema 4.1. *El determinante de una matriz de* 2×2 *satisface las tres condiciones siguientes:*

(a) *El determinante es una función lineal de cada renglón cuando el otro renglón permanece fijo; esto es,*

$$det \begin{pmatrix} cA_1 + A'_1 \\ A_2 \end{pmatrix} = c\ det \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} + det \begin{pmatrix} A'_1 \\ A_2 \end{pmatrix}$$

y

$$det \begin{pmatrix} A_1 \\ cA_2 + A'_2 \end{pmatrix} = c\ det \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} + det \begin{pmatrix} A_1 \\ A'_2 \end{pmatrix}$$

para todo escalar c *en* F.

(b) *Si* $A \in M_{2\times 2}(F)$ *tiene renglones idénticos, entonces* $det(A) = 0$.

(c) *Si* I *es la matriz identidad de* 2×2, *entonces* $det(I) = 1$.

DEMOSTRACIÓN.

(a) Sea $A_1 = (A_{11}\ A_{12})$, $A'_1 = (A'_{11}\ A'_{12})$ y $A_2 = (A_{21}\ A_{22})$;

entonces

$$\det\begin{pmatrix} cA_1 + A'_1 \\ A_2 \end{pmatrix} = \det\begin{pmatrix} cA_{11} + A'_{11} & cA_{12} + A'_{12} \\ A_{21} & A_{22} \end{pmatrix}$$
$$= (cA_{11} + A'_{11})A_{22} - (cA_{12} + A'_{12})A_{21}$$
$$= c(A_{11}A_{22} - A_{12}A_{21}) + A'_{11}A_{22} - A'_{12}A_{21})$$
$$= c \det\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} + \det\begin{pmatrix} A'_{11} & A'_{12} \\ A'_{21} & A'_{22} \end{pmatrix}$$
$$= c \det\begin{pmatrix} A_1 \\ A_2 \end{pmatrix} + \det\begin{pmatrix} A'_1 \\ A_2 \end{pmatrix}.$$

Un argumento semejante demuestra que el determinante es también una función lineal del segundo renglón.

(b) Si los renglones de A son idénticos, entonces A tiene la forma

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{11} & A_{12} \end{pmatrix}.$$

Así, $\det(A) = A_{11}A_{12} - A_{12}A_{11} = 0$.

(c) Puesto que

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$$\det(I) = 1 \cdot 1 - 0 \cdot 0 = 1. \quad ■$$

El siguiente resultado muestra que las tres propiedades mencionadas en el Teorema 4.1 caracterizan completamente al determinante tal como se definió anteriormente.

Teorema 4.2. *Sea* δ: $\mathsf{M}_{2\times 2}(\mathsf{F}) \to \mathsf{F}$ *una función cualquiera que tenga las tres propiedades siguientes:*

(a) δ *es una función lineal de cada renglón cuando el otro renglón se mantiene fijo.*

(b) *Si* $A \in \mathsf{M}_{2\times 2}(\mathsf{F})$ *tiene renglones idénticos, entonces* $\delta(A) = 0$.

(c) *Si* I *es la matriz identidad de* 2×2, *entonces* $\delta(I) = 1$.

Entonces $\delta = \det$; *esto es,* $\delta(A) = A_{11}A_{22} - A_{12}A_{21}$ *para toda* $A \in \mathsf{M}_{2\times 2}(\mathsf{F})$.

DEMOSTRACIÓN. Sea I la matriz identidad de 2×2 y sean

$$M_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \qquad M_2 = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{y} \quad M_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Obsérvese que $\delta(M_1) = \delta(M_2) = 0$ de acuerdo con la propiedad (b).

Primero demostraremos que $\delta(M_3) = -1$. Utilizando las propiedades (b) y (a) tenemos

$$\begin{aligned}
0 = \delta\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} &= \delta\begin{pmatrix} 1+0 & 0+1 \\ 1 & 1 \end{pmatrix} \\
&= \delta\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} + \delta\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \\
&= \delta\begin{pmatrix} 1 & 0 \\ 0+1 & 1+0 \end{pmatrix} + \delta\begin{pmatrix} 0 & 1 \\ 0+1 & 1+0 \end{pmatrix} \\
&= \delta\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \delta\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + \delta\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} + \delta\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\
&= \delta(I) + \delta(M_1) + \delta(M_2) + \delta(M_3) \\
&= 1 + 0 + 0 + \delta(M_3).
\end{aligned}$$

Luego, entonces $\delta(M_3) = -1$.

Ahora sea A un elemento cualquiera de $\mathsf{M}_{2\times 2}(F)$; entonces

$$\begin{aligned}
\delta(A) = \delta\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} &= \delta\begin{pmatrix} A_{11}+0 & 0+A_{12} \\ A_{21} & A_{22} \end{pmatrix} \\
&= \delta\begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix} + \delta\begin{pmatrix} 0 & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \\
&= \delta\begin{pmatrix} A_{11} & 0 \\ 0+A_{21} & A_{22}+0 \end{pmatrix} + \delta\begin{pmatrix} 0 & A_{12} \\ 0+A_{21} & A_{22}+0 \end{pmatrix} \\
&= \delta\begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} \end{pmatrix} + \delta\begin{pmatrix} A_{11} & 0 \\ A_{21} & 0 \end{pmatrix} + \delta\begin{pmatrix} 0 & A_{12} \\ 0 & A_{22} \end{pmatrix} + \delta\begin{pmatrix} 0 & A_{12} \\ A_{21} & 0 \end{pmatrix} \\
&= A_{11}A_{22}\cdot\delta\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + A_{11}A_{21}\cdot\delta\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + A_{12}A_{22}\cdot\delta\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \\
&\quad + A_{12}A_{21}\cdot\delta\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\
&= A_{11}A_{22}\cdot\delta(I) + A_{11}A_{21}\cdot\delta(M_1) + A_{12}A_{22}\cdot\delta(M_2) + A_{12}A_{21}\cdot\delta(M_3) \\
&= A_{11}A_{22}(1) + A_{11}A_{21}(0) + A_{12}A_{22}(0) + A_{12}A_{21}(-1) \\
&= A_{11}A_{22} - A_{12}A_{21} = \det(A).
\end{aligned}$$

Y, por tanto, $\delta = \det$. ■

Motivados por esta caracterización del determinante de una matriz de 2×2, en la Sección 4.2 definiremos un determinante en $\mathsf{M}_{n\times n}(F)$ como una función que posee las tres propiedades del Teorema 4.1. Pero primero veremos esta propiedad de unicidad para estudiar el significado geométrico del determinante de una matriz de 2×2. En particular, encon-

traremos que el signo del determinante es de importancia geométrica en el estudio de la orientación.

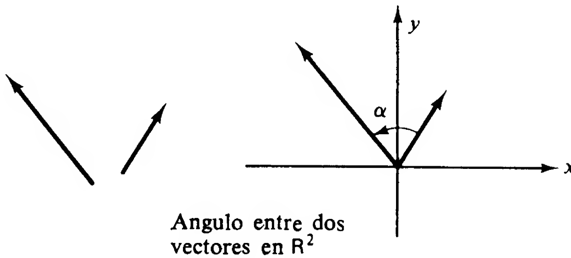
Al hablar del *ángulo entre dos vectores* en R^2, se entiende que hablamos del ángulo θ, tal que $0 \leq \theta < \pi$, formado por los vectores de la misma magnitud y dimensión que los vectores dados pero que parten del origen. (Véase Fig. 4.1.) Dados tres vectores u, v y w que parten del mismo punto, se dice que v está ubicado *entre* u y w si el ángulo entre u y w es igual a la suma de los ángulos entre u y v y entre v y w. (Véase Fig. 4.2.)

Dada una base ordenada $\beta = \{u, v\}$ para R^2, donde $u = (a_1, a_2)$ y $v = (b_1, b_2)$ denotamos por

$$\det\begin{pmatrix} u \\ v \end{pmatrix}$$

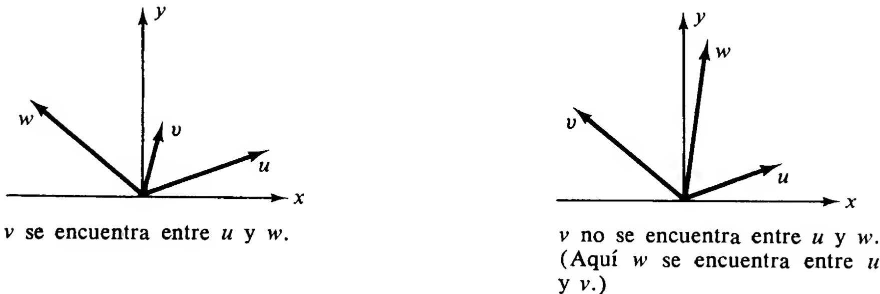
al escalar

$$\det\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix},$$



Angulo entre dos vectores en R^2

figura 4.1



v se encuentra entre u y w.

v no se encuentra entre u y w. (Aquí w se encuentra entre u y v.)

figura 4.2

y definimos la *orientación* de β como el número real

$$O\begin{pmatrix} u \\ v \end{pmatrix} = \frac{\det\begin{pmatrix} u \\ v \end{pmatrix}}{\left|\det\begin{pmatrix} u \\ v \end{pmatrix}\right|}.$$

(Se deduce del Ejercicio 10 que el denominador no es cero.) Claramente

$$O\begin{pmatrix} u \\ v \end{pmatrix} = \pm 1.$$

Obsérvese que

$$O\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = 1 \quad \text{y} \quad O\begin{pmatrix} e_1 \\ -e_2 \end{pmatrix} = -1.$$

En general (véase el Ejercicio 11),

$$O\begin{pmatrix} u \\ v \end{pmatrix} = 1$$

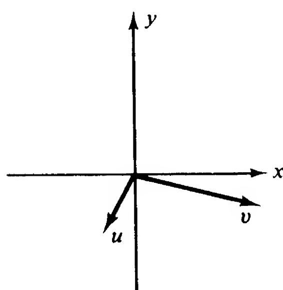
si y sólo si la base ordenada $\{u, v\}$ forma un sistema coordenado derecho, y

$$O\begin{pmatrix} u \\ v \end{pmatrix} = -1$$

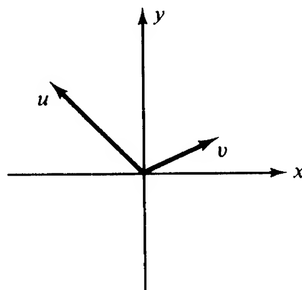
si y sólo si $\{u, v\}$ forma un sistema coordenado izquierdo. (Recuérdese que un sistema coordenado $\{u, v\}$ es derecho si u puede hacerse coincidir con v haciéndolo girar en contra de las manecillas del reloj un ángulo θ tal que $0 < \theta < \pi$; de lo contrario $\{u, v\}$ es un sistema coordenado izquierdo. Véase Fig. 4.3.) Por conveniencia, definimos

$$O\begin{pmatrix} u \\ v \end{pmatrix} = 0$$

si $\{u, v\}$ es linealmente dependiente.



Un sistema coordenado derecho

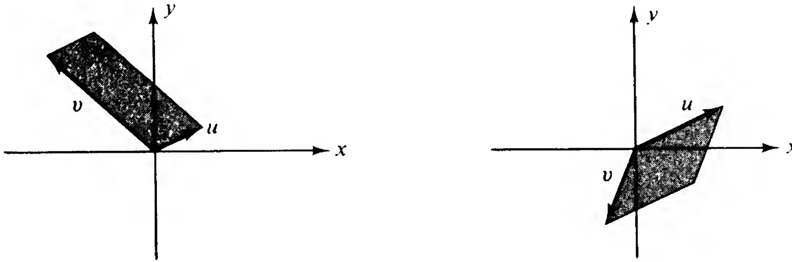


Un sistema coordenado izquierdo

figura 4.3

Cualquier conjunto ordenado $\{u, v\}$ en R^2 determina de la siguiente manera un paralelogramo. Considerando a u y a v como flechas que parten del origen de R^2, llamamos al paralelogramo que tiene a u y a v

como lados adyacentes el *paralelogramo determinado por* u *y* v (véase Fig. 4.4).



Paralelogramos determinados por u y v

figura 4.4

Obsérvese que si el conjunto $\{u, v\}$ es linealmente dependiente, es decir, si u y v son paralelos, el "paralelogramo" determinado por u y v es en realidad un segmento de recta que podemos considerar como un paralelogramo degenerado cuya área es cero.

Existe una relación interesante entre

$$\mathrm{A}\begin{pmatrix} u \\ v \end{pmatrix},$$

el área del paralelogramo determinado por u y v, y

$$\det\begin{pmatrix} u \\ v \end{pmatrix},$$

que ahora procederemos a investigar. Sin embargo, obsérvese primero que como

$$\det\begin{pmatrix} u \\ v \end{pmatrix}$$

puede ser negativo, no podemos esperar que

$$\mathrm{A}\begin{pmatrix} u \\ v \end{pmatrix} = \det\begin{pmatrix} u \\ v \end{pmatrix}.$$

Pero podemos demostrar que

$$\mathrm{A}\begin{pmatrix} u \\ v \end{pmatrix} = \mathrm{O}\begin{pmatrix} u \\ v \end{pmatrix} \cdot \det\begin{pmatrix} u \\ v \end{pmatrix},$$

de donde se tiene que

$$\mathrm{A}\begin{pmatrix} u \\ v \end{pmatrix} = \left| \det\begin{pmatrix} u \\ v \end{pmatrix} \right|.$$

Con el argumento de que

$$A\binom{u}{v} = O\binom{u}{v} \cdot \det\binom{u}{v},$$

utilizaremos una técnica que, aun cuando algo indirecta, podrá ser generalizada para R^n. Puesto que

$$O\binom{u}{v} = \pm 1,$$

podemos multiplicar ambos lados de la ecuación deseada por

$$O\binom{u}{v}$$

para obtener la forma equivalente

$$O\binom{u}{v} \cdot A\binom{u}{v} = \det\binom{u}{v}.$$

Estableceremos esta ecuación verificando que las tres condiciones del Teorema 4.2 se satisfacen por la función

$$\delta\binom{u}{v} = O\binom{u}{v} \cdot A\binom{u}{v}.$$

(a) Principiaremos demostrando que

$$\delta\binom{u}{\lambda v} = \lambda \cdot \delta\binom{u}{v}.$$

Obsérvese que esta conclusión es inmediata si $\lambda = 0$ puesto que

$$\delta\binom{u}{\lambda v} = O\binom{u}{0} \cdot A\binom{u}{0} = 0.$$

Supóngase entonces que $\lambda \neq 0$. Considerando a λv como la base del paralelogramo determinado por u y λv, vemos que

$$A\binom{u}{\lambda v} = \text{base} \times \text{altura} = |\lambda|(\text{longitud de } v)(\text{altura}) = |\lambda| \cdot A\binom{u}{v}$$

puesto que la altura h del paralelogramo determinado por u y λv, es la misma que la del paralelogramo determinado por u y v. (Véase Fig. 4.5.) De aquí que

$$\delta\binom{u}{\lambda v} = O\binom{u}{\lambda v} \cdot A\binom{u}{\lambda v} = \left[\frac{\lambda}{|\lambda|} \cdot O\binom{u}{v}\right]\left[|\lambda| \cdot A\binom{u}{v}\right]$$

$$= \lambda \cdot O\binom{u}{v} \cdot A\binom{u}{v} = \lambda \cdot \delta\binom{u}{v}.$$

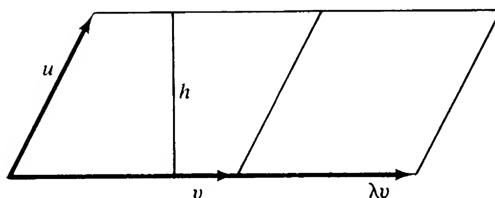


figura 4.5

Un argumento semejante muestra que

$$\delta\begin{pmatrix} \lambda u \\ u \end{pmatrix} = \lambda \cdot \delta\begin{pmatrix} u \\ v \end{pmatrix}.$$

Demostraremos a continuación que

$$\delta\begin{pmatrix} u \\ au + bw \end{pmatrix} = b \cdot \delta\begin{pmatrix} u \\ w \end{pmatrix}$$

para toda u, $w \in R^2$ y todo número real a y b. Obsérvese que debido a que los paralelogramos determinados por u y w y por u y $u + w$ tienen una base común u y la misma altura (véase Fig. 4.6),

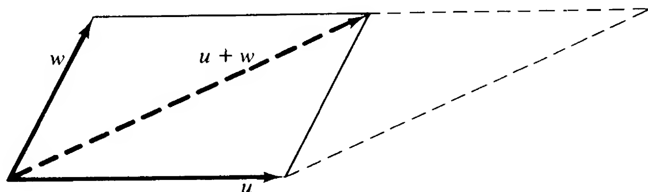


figura 4.6

$$A\begin{pmatrix} u \\ w \end{pmatrix} = A\begin{pmatrix} u \\ u + w \end{pmatrix}.$$

Si $a = 0$, entonces

$$\delta\begin{pmatrix} u \\ au + bw \end{pmatrix} = \delta\begin{pmatrix} u \\ bw \end{pmatrix} = b \cdot \delta\begin{pmatrix} u \\ w \end{pmatrix}$$

de acuerdo con el primer párrafo de la parte (a). De lo contrario, si $a \neq 0$, entonces

$$\delta\begin{pmatrix} u \\ au + bw \end{pmatrix} = a \cdot \delta\begin{pmatrix} u \\ u + \frac{b}{a}w \end{pmatrix} = a \cdot \delta\begin{pmatrix} u \\ \frac{b}{a}w \end{pmatrix} = b \cdot \delta\begin{pmatrix} u \\ w \end{pmatrix}.$$

Y así se obtiene la condición deseada en ambos casos.

Ahora podremos demostrar que

$$\delta\begin{pmatrix} u \\ v_1 + v_2 \end{pmatrix} = \delta\begin{pmatrix} u \\ v_1 \end{pmatrix} + \delta\begin{pmatrix} u \\ v_2 \end{pmatrix}$$

para toda u, v_1, $v_2 \in \mathsf{R}^2$. Como el resultado es inmediato si $u = 0$, supondremos que $u \neq 0$. Tómese cualquier vector $w \in \mathsf{R}^2$ tal que $\{u, w\}$ sea linealmente independiente. Entonces, para vectores cualesquiera v_1, $v_2 \in \mathsf{R}^2$, existen escalares a_i y b_i tales que $v_i = a_i u + b_i w (i = 1, 2)$. Entonces

$$\delta\begin{pmatrix} u \\ v_1 + v_2 \end{pmatrix} = \delta\begin{pmatrix} u \\ (a_1 + a_2)u + (b_1 + b_2)w \end{pmatrix} = (b_1 + b_2)\delta\begin{pmatrix} u \\ w \end{pmatrix}$$

$$= \delta\begin{pmatrix} u \\ a_1 u + b_1 w \end{pmatrix} + \delta\begin{pmatrix} u \\ a_2 u + b_2 w \end{pmatrix} = \delta\begin{pmatrix} u \\ v_1 \end{pmatrix} + \delta\begin{pmatrix} u \\ v_2 \end{pmatrix}$$

Un argumento semejante muestra que

$$\delta\begin{pmatrix} u_1 + u_2 \\ v \end{pmatrix} = \delta\begin{pmatrix} u_1 \\ v \end{pmatrix} + \delta\begin{pmatrix} u_2 \\ v \end{pmatrix}$$

para todo u_1, u_2, $v \in \mathsf{R}^2$.

(b) Como

$$\mathrm{A}\begin{pmatrix} u \\ u \end{pmatrix} = 0, \qquad \delta\begin{pmatrix} u \\ u \end{pmatrix} = 0 = \mathrm{O}\begin{pmatrix} u \\ u \end{pmatrix} \cdot \mathrm{A}\begin{pmatrix} u \\ u \end{pmatrix}$$

para toda $u \in \mathsf{R}^2$.

(c) Como el paralelogramo definido por e_1 y e_2 es el cuadrado unitario,

$$\delta\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = 1 = \mathrm{O}\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \cdot \mathrm{A}\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}.$$

Por tanto, δ satisface las tres condiciones del Teorema 4.2 y $\delta = \det$. Así, el área del paralelogramo determinado por u y v es igual a

$$\mathrm{O}\begin{pmatrix} u \\ v \end{pmatrix} \cdot \det\begin{pmatrix} u \\ v \end{pmatrix}.$$

Entonces, por ejemplo, se ve que el área del paralelogramo determinado por $u = (-1, 5)$ y $v = (4, -2)$ es

$$\left|\det\begin{pmatrix} u \\ v \end{pmatrix}\right| = \left|\det\begin{pmatrix} -1 & 5 \\ 4 & -2 \end{pmatrix}\right| = 18.$$

EJERCICIOS

1. Decir si las siguientes afirmaciones son falsas o verdaderas.

   (a) El determinante de una matriz de 2×2 es una función lineal de cada renglón de la matriz cuando el otro renglón se mantiene fijo.

   (b) Si I es la matriz identidad de 2×2, entonces $\det(I) = 0$.

(c) Si ambos renglones de una matriz A de 2×2 son idénticos, entonces $\det(A) = 0$.

(d) Si u y v son vectores en R^2 que parten del origen, entonces el área del paralelogramo que tiene a u y v como lados adyacentes es

$$\det\begin{pmatrix} u \\ v \end{pmatrix}.$$

(e) Un sistema coordenado $\{u, v\}$ es derecho si y sólo si su orientación es 1.

(f) El determinante es una transformación lineal de $\mathsf{M}_{2\times 2}(F)$ en F.

2. Calcular los determinantes de los siguientes elementos de $\mathsf{M}_{2\times 2}(R)$:

(a) $\begin{pmatrix} 6 & -3 \\ 2 & 4 \end{pmatrix}$ (b) $\begin{pmatrix} -5 & 2 \\ 6 & 1 \end{pmatrix}$ (c) $\begin{pmatrix} 8 & 0 \\ 3 & -1 \end{pmatrix}$

3. Calcular los determinantes de los siguientes elementos de $\mathsf{M}_{2\times 2}(C)$:

(a) $\begin{pmatrix} -1+i & 1-4i \\ 3+2i & 2-3i \end{pmatrix}$ (b) $\begin{pmatrix} 5-2i & 6+4i \\ -3+i & 7i \end{pmatrix}$ (c) $\begin{pmatrix} 2i & 3 \\ 4 & 6i \end{pmatrix}$

4. Para cada uno de los siguientes pares de vectores u y v en R^2, calcule el área del paralelogramo determinado por u y v.

(a) $u = (3, -2)$ y $v = (2, 5)$
(b) $u = (1, 3)$ y $v = (-3, 1)$
(c) $u = (4, -1)$ y $v = (-6, -2)$
(d) $u = (3, 4)$ y $v = (2, -6)$

5. Demostrar que si B es la matriz obtenida al intercambiar los renglones de una matriz A de 2×2, entonces $\det(B) = -\det(A)$.

6. Demostrar que para cualquier $A \in \mathsf{M}_{2\times 2}(F)$, $\det(A^t) = \det(A)$.

7. Demostrar que si A es una matriz triangular de 2×2, entonces el determinante de A es igual al producto de los elementos de A situados en la diagonal.

8. Demostrar que para cualesquiera A, $B \in \mathsf{M}_{2\times 2}(F)$, $\det(AB) = \det(A) \cdot \det(B)$.

9. La *adjunta clásica* de una matriz A de 2×2 es la matriz

$$\operatorname{adj} A = \begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix}.$$

Demostrar que la adjunta clásica de una matriz posee las siguientes propiedades:

(a) $(\operatorname{adj} A)A = A(\operatorname{adj} A) = [\det(A)]I$.
(b) $\det(\operatorname{adj} A) = \det(A)$.
(c) $\operatorname{adj} A^t = (\operatorname{adj} A)^t$.

10. Utilizando al Ejercicio 9(a), demostrar que una matriz A de 2×2 es invertible si y sólo si $\det(A) \neq 0$, y que en este caso $A^{-1} = [\det(A)]^{-1} \times (\operatorname{adj} A)$.

11. Demostrar que

$$O\begin{pmatrix} u \\ v \end{pmatrix} = 1$$

si y sólo si la base ordenada $\{u, v\}$ para R^2 forma un sistema coordenado derecho. *Sugerencia:* Recordar la definición de una rotación dada en el Ejemplo 5 de la Sección 2.1.

4.2 DETERMINANTES DE ORDEN *n*

Hemos visto en el Teorema 4.2 que el determinante de una matriz de 2×2 se caracteriza totalmente por tres propiedades. Definiremos pronto el determinante de una matriz de $n \times n$ en términos de esas propiedades, pero primero necesitaremos de algunos resultados preliminares. Para comenzar definiremos la primera de las condiciones que caracterizaron al determinante de una matriz de 2×2.

Definición. *Una función* δ: $M_{n\times n}(F) \to F$ *se dice que es una* función n-lineal *si* δ *es una función lineal de cada renglón de una matriz de* $n \times n$ *cuando los restantes* $n - 1$ *renglones se mantienen fijos, esto es, si*

$$\delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ cA_i + A_i' \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} = c\cdot\delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} + \delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i' \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix}, \text{ para } i = 1, 2, \ldots, n$$

siempre y cuando

$$\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ cA_i + A_i' \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix}$$

sea un elemento de $M_{n\times n}(F)$.

Ejemplo 2. El Teorema 4.1 muestra que det: $M_{2\times 2}(F) \to F$ definido por $\det(A) = A_{11}A_{22} - A_{12}A_{21}$ es una función 2-lineal.

Ejemplo 3. La función δ: $M_{n\times n}(F) \to F$ definida por $\delta(A) = 0$ para toda $A \in M_{n\times n}(F)$ es una función n-lineal.

Ejemplo 4. La función δ: $M_{n\times n}(F) \to F$ definida mediante $\delta(A) = A_{1j}A_{2j} \ldots A_{nj}$ (esto es $\delta(A)$ es igual al producto de todos los elementos de la j-ésima columna de A) es una función n-lineal para cada $j(1 \leq j \leq n)$ puesto que

$$\delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ cA_i + A_i' \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} = A_{1j} \cdots A_{(i-1)j}(cA_{ij} + A'_{ij})A_{(i+1)j} \cdots A_{nj}$$

$$= c(A_{1j} \cdots A_{ij} \cdots A_{nj}) + (A_{1j} \cdots A_{(i-1)j}A'_{ij}A_{(i+1)j} \cdots A_{nj})$$

$$= c\cdot\delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} + \delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i' \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix}.$$

Ejemplo 5. La función δ: $M_{n\times n}(F) \to F$ definida mediante $\delta(A) = A_{11}A_{22} \ldots A_{nn}$ (esto es, $\delta(A)$ es igual al producto de los elementos de A ubicados sobre la diagonal) es una función n-lineal.

Ejemplo 6. La función δ: $M_{n\times n}(F) \to F$ definida por $\delta(A) = \operatorname{tr}(A)$ no es una función n-lineal.

Nuestro siguiente resultado muestra que las funciones n-lineales pueden combinarse para producir otras funciones n-lineales.

Teorema 4.3. *Una combinación lineal de dos funciones* n-*lineales es una función* n-*lineal (donde la suma y la multiplicación por escalares son como se definieron en el Ejemplo 3 de la Sección 1.2).*

DEMOSTRACIÓN. Sean δ_1 y δ_2 funciones n-lineales y sean a y b escalares. Si δ es la combinación lineal $\delta = a\delta_1 + b\delta_2$, entonces

$$\delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ cA_i + A_i' \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} = a\cdot\delta_1\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ cA_i + A_i' \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} + b\cdot\delta_2\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ cA_i + A_i' \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix}$$

$$= a\left[c\cdot\delta_1\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} + \delta_1\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i' \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix}\right] + b\left[c\cdot\delta_2\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} + \delta_2\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i' \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix}\right]$$

$$= c\left[a\cdot\delta_1\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} + b\cdot\delta_2\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix}\right] + \left[a\cdot\delta_1\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i' \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} + b\cdot\delta_2\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i' \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix}\right]$$

$$= c\cdot\delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} + \delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i' \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} \qquad \text{para toda } i,\ 1 \leq i \leq n.$$

Entonces δ es una función n-lineal. ■

Corolario. *Cualquier combinación lineal de funciones* n-*lineales es una función* n-*lineal.*

DEMOSTRACIÓN. Ejercicio.

La siguiente definición hace mención de la segunda de las tres propiedades que caracterizaron al determinante de una matriz de 2×2.

Definición. *Se dice que una función* δ n-*lineal es alternante si* $\delta(A) = 0$ *siempre que dos renglones adyacentes sean idénticos.*

Ejemplo 7. De las tres funciones n-lineales dadas en los Ejemplos 3, 4 y 5 sólo la primera de ellas es alternante.

El siguiente resultado muestra que la definición anterior es más poderosa de lo que parece. En particular, no hay necesidad de que los renglones en la definición se supongan adyacentes.

Teorema 4.4. *Sea* δ: $M_{n\times n}(F) \to F$ *una función* n-*lineal alternante. Entonces son ciertas las siguientes expresiones:*

(a) *Si* B *se obtiene al intercambiar cualquier par de renglones en una matriz* A *de* $n \times n$, *entonces* $\delta(B) = -\delta(A)$.
(b) *Si dos renglones de una matriz de* $n \times n$ *son idénticos, entonces* $\delta(A) = 0$.

DEMOSTRACIÓN. Demostraremos primero que si B se obtiene al intercambiar cualquier par de renglones adyacentes de A entonces $\delta(B) = -\delta(A)$. Supóngase que B se obtiene al intercambiar los renglones i e $i + 1$ de

$$A = \begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ A_{i+1} \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix}; \text{ entonces } \quad B = \begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_{i+1} \\ A_i \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix}.$$

Ahora bien

$$0 = \delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i + A_{i+1} \\ A_i + A_{i+1} \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} = \delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ A_i + A_{i+1} \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} + \delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_{i+1} \\ A_i + A_{i+1} \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix}$$

$$= \delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ A_i \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} + \delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ A_{i+1} \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} + \delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_{i+1} \\ A_i \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} + \delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_{i+1} \\ A_{i+1} \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix}$$

$$= 0 + \delta(A) + \delta(B) + 0$$

porque δ es una función n-lineal alternante. Así, $\delta(B) = -\delta(A)$.

Ahora supóngase que B se obtiene de A intercambiando los renglones i y j donde $i < j$. Comenzando con los renglones i e $i+1$, intercambiamos sucesivamente los renglones de A hasta que éstos tienen el orden siguiente:

$$A_1, \ldots, A_{i-1}, A_{i+1}, \ldots, A_j, A_i, A_{j+1}, \ldots, A_n.$$

En total, se requieren $j - i$ intercambios para producir este orden. Ahora intercámbiese sucesivamente A_j con el renglón anterior hasta que los renglones tengan el orden siguiente:

$$A_1, \ldots, A_{i-1}, A_j, A_{i+1}, \ldots, A_{j-1}, A_i, A_{j+1}, \ldots, A_n.$$

Este proceso requiere de $j - i - 1$ intercambios de renglones adyacentes y produce la matriz B. De aquí, en virtud del primer párrafo de la demostración, vemos que

$$\delta(B) = (-1)^{j-i}(-1)^{j-i-1}\,\delta(A) = (-1)^{2(j-i)-1}\,\delta(A) = -\delta(A).$$

Resta demostrar que si dos renglones de A son idénticos, por ejemplo i y j $(i < j)$, entonces $\delta(A) = 0$. Si $j = i + 1$, entonces dos renglones adyacentes de A son idénticos y por hipótesis $\delta(A) = 0$. Si $j > i + 1$, intercámbiense los renglones $i + 1$ y j para obtener una matriz B con dos renglones adyacentes iguales. Entonces $\delta(B) = 0$, pero como $\delta(B) = -\delta(A)$ de acuerdo con el segundo párrafo de la demostración, se tiene que $\delta(A) = 0$. De este modo δ satisface las condiciones (a) y (b). ■

Estamos preparados ahora para definir un determinante en $\mathsf{M}_{n\times n}(F)$. Obsérvese que el determinante se define en términos de las tres propiedades del Teorema 4.2 que caracterizan al determinante de una matriz de 2×2.

Definición. *Un* determinante *en* $\mathsf{M}_{n\times n}(F)$ *es una función alternante* n*-lineal* $\delta\colon \mathsf{M}_{n\times n}(F) \to F$ *tal que* $\delta(I) = 1$.

Un ejemplo sencillo de determinante puede darse en $\mathsf{M}_{1\times 1}(F)$, para la función $\delta\colon \mathsf{M}_{1\times 1}(F) \to F$ definida por $\delta(A) = A_{11}$ (el único elemento

de A), que claramente satisface los requisitos de esta definición. Mas aún, el Teorema 4.1 muestra que al definir el determinante de una matriz A de 2×2 como $A_{11}A_{22} - A_{12}A_{21}$ se obtiene un determinante en $\mathsf{M}_{2\times 2}(F)$ en el sentido de la definición anterior. Nuestro siguiente resultado nos autoriza a definir un determinante en $\mathsf{M}_{n\times n}(F)$ por inducción para cualquier $n \geq 3$.

Teorema 4.5. *Sea* δ *una función* n*-lineal alternante en* $\mathsf{M}_{n\times n}(F)$. *Para cada matriz* A *de* $(n+1)\times(n+1)$ *y para cada* $j(1 \leq j \leq n+1)$, *se define*

$$\epsilon_j(A) = \sum_{i=1}^{n+1}(-1)^{i+j}A_{ij}\cdot\delta(\tilde{A}_{ij}),$$

donde $\tilde{A}_{ij}$ *es la matriz de* $n \times n$ *obtenida a partir de* A *eliminando el i-ésimo renglón y la j-ésima columna. Entonces* ϵ_j *es una función* $(n+1)$*-lineal alternante en las matrices de* $(n+1)\times(n+1)$ *con elementos de* F.

DEMOSTRACIÓN. Como $\tilde{A}_{ij}$ se obtiene a partir de A al suprimir el i-ésimo renglón y la j-ésima columna, $\delta(\tilde{A}_{ij})$ es independiente del i-ésimo renglón de A. Entonces, como δ es una función n-lineal $\delta(\tilde{A}_{ij})$ es una función lineal de cada renglón de A a excepción del renglón i. Por tanto $A_{ij}\cdot\delta(\tilde{A}_{ij})$ es una función $(n+1)$-lineal de las matrices de $(n+1)\times(n+1)$ con elementos de F. Entonces, como

$$\epsilon_j(A) = \sum_{i=1}^{n+1}(-1)^{i+j}A_{ij}\cdot\delta(\tilde{A}_{ij})$$

es una combinación lineal de las funciones $(n+1)$-lineales $A_{ij}\cdot\delta(\tilde{A}_{ij})$, ϵ_j es una función $(n+1)$-lineal en virtud del corolario del Teorema 4.3.

Demostraremos ahora que ϵ_j es alternante. Supóngase que A es una matriz de $(n+1)\times(n+1)$ en la que los renglones k y $k+1$ son idénticos. Entonces $\tilde{A}_{ij}$ tiene dos renglones idénticos siempre que $i \neq k$ e $i \neq k+1$. Así $\delta(\tilde{A}_{ij}) = 0$ siempre que $i \neq k$ e $i \neq k+1$ y entonces

$$\epsilon_j(A) = (-1)^{k+j}A_{kj}\cdot\delta(\tilde{A}_{kj}) + (-1)^{(k+1)+j}A_{(k+1)j}\cdot\delta(\tilde{A}_{(k+1)j}).$$

Pero como los renglones k y $k+1$ de A son iguales, $A_{kj} = A_{(k+1)j}$ y $\tilde{A}_{kj} = \tilde{A}_{(k+1)j}$. De aquí que $\epsilon_j(A) = 0$, lo que demuestra que ϵ_j es alternante. ■

Corolario 1. *Sean* δ *y* ϵ_j *como en el enunciado del Teorema 4.5. Si* δ *es un determinante en* $\mathsf{M}_{n\times n}(F)$, *entonces* ϵ_j *es un determinante en las matrices de* $(n+1)\times(n+1)$ *con elementos de* F.

DEMOSTRACIÓN. Sea I la matriz identidad de $(n+1)\times(n+1)$ y sea $\tilde{I}_{jj}$ la matriz de $n \times n$ obtenida a partir de I al suprimir el renglón i y

la columna j. Entonces $\tilde{I}_{jj}$ es la matriz identidad de $n \times n$. Como $I_{ij} = 0$ si $i \neq j$ e $I_{jj} = 1$, tenemos

$$\begin{aligned} \epsilon_j(I) &= \sum_{i=1}^{n+1} (-1)^{i+j} I_{ij} \cdot \delta(\tilde{I}_{ij}) = (-1)^{j+j} \cdot \delta(\tilde{I}_{jj}) \\ &= \delta(\tilde{I}_{jj}) = 1 \end{aligned}$$

debido a que δ es un determinante en $\mathsf{M}_{n\times n}(F)$. Entonces ϵ_j es un determinante en las matrices de $(n+1) \times (n+1)$ con elementos de F. ■

Corolario 2. *Existe un determinante en* $\mathsf{M}_{n\times n}(F)$ *para todo entero positivo* n.

DEMOSTRACIÓN. La demostración será por inducción sobre n. Si $n = 1$, la función det: $\mathsf{M}_{1\times 1}(F) \to F$ definida mediante $\det(A) = A_{11}$ es un determinante en $\mathsf{M}_{1\times 1}(F)$. Supóngase que existe un determinante δ en $\mathsf{M}_{n\times n}(F)$. Entonces para alguna $j(1 \leq j \leq n+1)$, la función ϵ_j definida en el Teorema 4.5 es un determinante en las matrices de $(n+1) \times (n+1)$ con elementos de F. Esto completa la inducción. ■

Definiciones. *Si* δ *es un determinante en* $\mathsf{M}_{n\times n}(F)$, *entonces el determinante*

$$\epsilon_j(A) = \sum_{i=1}^{n+1} (-1)^{i+j} A_{ij} \cdot \delta(\tilde{A}_{ij})$$

definido en el Teorema 4.5 se llama la expansión de A a lo largo de la *j*-ésima columna. *El escalar* $(-1)^{i+j} \cdot \delta(\tilde{A}_{ij})$ *se llama el* cofactor de A_{ij} *(con respecto al determinante* δ*).*

Ejemplo 8. Sea A el siguiente elemento de $\mathsf{M}_{3\times 3}(F)$:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Los cofactores de A_{12}, A_{22} y A_{32} son, respectivamente,

$$(-1)^{1+2} \det\begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix} = (-1)(4 \cdot 9 - 6 \cdot 7) = 6,$$

$$(-1)^{2+2} \det\begin{pmatrix} 1 & 3 \\ 7 & 9 \end{pmatrix} = 1(1 \cdot 9 - 3 \cdot 7) = -12,$$

$$(-1)^{3+2} \det\begin{pmatrix} 1 & 3 \\ 4 & 6 \end{pmatrix} = (-1)(1 \cdot 6 - 3 \cdot 4) = 6.$$

Por tanto, la expansión de A sobre la segunda columna es

$$\begin{aligned} \epsilon_2(A) &= A_{12}(6) + A_{22}(-12) + A_{32}(6) \\ &= 2 \cdot 6 + 5(-12) + 8 \cdot 6 = 0. \end{aligned}$$

De la misma manera, los cofactores de A_{13}, A_{23} y A_{33} son, respectivamente,

$$(-1)^{1+3}\det\begin{pmatrix}4 & 5\\ 7 & 8\end{pmatrix} = 1(4\cdot 8 - 5\cdot 7) = -3,$$

$$(-1)^{2+3}\det\begin{pmatrix}1 & 2\\ 7 & 8\end{pmatrix} = (-1)(1\cdot 8 - 2\cdot 7) = 6,$$

$$(-1)^{3+3}\det\begin{pmatrix}1 & 2\\ 4 & 5\end{pmatrix} = 1(1\cdot 5 - 2\cdot 4) = -3.$$

De aquí que la expansión de A a lo largo de la tercera columna es

$$\epsilon_3(A) = A_{13}(-3) + A_{23}(6) + A_{33}(-3) = 3(-3) + 6(6) + \\ + 9(-3) = 0.$$

Veremos en el Teorema 4.9 que la igualdad de $\epsilon_2(A)$ y $\epsilon_3(A)$ en el Ejemplo 8 no es coincidencia. De hecho, veremos que existe exactamente un determinante en $\mathsf{M}_{n\times n}(F)$.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.
   (a) Un determinante en $\mathsf{M}_{n\times n}(F)$ es una función lineal de cada renglón de una matriz de $n \times n$ con elementos de F cuando el resto de los $n - 1$ renglones permanece fijo.
   (b) Si δ es un determinante y cualquier par de renglones de A son idénticos, entonces $\delta(A) = 0$.
   (c) Sea δ un determinante. Si B es una matriz obtenida a partir de A intercambiando dos renglones cualesquiera entonces $\delta(A) = \delta(B)$.
   (d) La función $\delta\colon \mathsf{M}_{n\times n}(F) \to F$ definida por $\delta(A) = 0$ para toda $A \in \mathsf{M}_{n\times m}(F)$ es un determinante en $\mathsf{M}_{n\times n}(F)$.
   (e) Para cualquier $n \geq 2$ existe un determinante en $\mathsf{M}_{n\times n}(F)$.
   (f) Cualquier determinante $\delta\colon \mathsf{M}_{n\times n}(F) \to F$ es lineal.

2. Verificar que si A es la matriz de 3×3 del Ejemplo 8, entonces la expansión de A a lo largo de la primera columna es igual a cero.

3. Evaluar el determinante de cada una de las siguientes matrices por expansión sobre la segunda y la tercera columna. (Cada matriz es un elemento de $\mathsf{M}_{3\times 3}(C)$.)

   (a) $\begin{pmatrix}-3 & 2 & 5\\ 1 & 0 & -1\\ 4 & -6 & 7\end{pmatrix}$ (b) $\begin{pmatrix}8 & -4 & 0\\ 0 & 6 & -3\\ -1 & 5 & 2\end{pmatrix}$

(c) $\begin{pmatrix} 1 & 2 & -5 \\ 6 & -4 & 3 \\ 0 & 1 & 1 \end{pmatrix}$ (d) $\begin{pmatrix} 1+i & -1 & 0 \\ 2 & 3i & 4i \\ 0 & 2-i & -1+2i \end{pmatrix}$

4. ¿Cuáles de las siguientes funciones $\delta\colon \mathsf{M}_{3\times 3}(F) \to F$ son funciones 3-lineales? Justificar cada respuesta.

(a) $\delta(A) = c$, donde c es cualquier escalar no nulo
(b) $\delta(A) = A_{22}$
(c) $\delta(A) = A_{11}A_{23}A_{32}$
(d) $\delta(A) = A_{11}A_{21}A_{32}$
(e) $\delta(A) = A_{11}A_{31}A_{32}$
(f) $\delta(A) = A_{11}^2A_{22}^2A_{33}^2$
(g) $\delta(A) = A_{11}A_{22}A_{33} - A_{11}A_{21}A_{32}$

5. (a) Determinar todas las funciones 1-lineales $\delta\colon \mathsf{M}_{1\times 1}(F) \to F$.
(b) Determinar todos los determinantes en $\mathsf{M}_{1\times 1}(F)$.

6. Demostrar la igualdad de las tres funciones $\epsilon_j\colon \mathsf{M}_{3\times 3}(F) \to F (j = 1, 2, 3)$ definidas en el Teorema 4.5 para toda $A \in \mathsf{M}_{3\times 3}(F)$ por

$$\epsilon_j(A) = \sum_{i=1}^{3} (-1)^{i+j} A_{ij} \cdot \det(\tilde{A}_{ij}),$$

donde $\tilde{A}_{ij}$ es la matriz de 2×2 obtenida a partir de A suprimiendo el renglón i y la columna j y det denota el determinante único en $\mathsf{M}_{2\times 2}(F)$.

7. Demostrar que el determinante único en $\mathsf{M}_{2\times 2}(F)$ es una función 2-lineal de las columnas de una matriz de 2×2 y que el determinante de una matriz de 2×2, en la que ambas columnas son idénticas, es cero.

8. La demostración del Teorema 4.2 muestra que si δ es una función 2-lineal $\delta\colon \mathsf{M}_{2\times 2}(F) \to F$, entonces

$$\delta(A) = A_{11}A_{22} \cdot \delta(I) + A_{11}A_{21} \cdot \delta(M_1) + A_{12}A_{22} \cdot \delta(M_2) + A_{12}A_{21} \cdot \delta(A_3),$$

donde I, M_1, M_2 y M_3 son como en la demostración del teorema. Demostrar que para escalares cualesquiera $a, b, c, d \in F$ la función

$$\epsilon(A) = A_{11}A_{22}a + A_{11}A_{21}b + A_{12}A_{22}c + A_{12}A_{21}d$$

es 2-lineal. Entonces $\delta'\colon \mathsf{M}_{2\times 2}(F) \to F$ es una función 2-lineal si y sólo si es de la forma anterior para algunos escalares a, b, c y d.

9. Demostrar que si F no es un campo de característica dos (tal como se define en el Apéndice D), entonces la condición (a) del Teorema 4.4 im-

plica la condición (b) del mismo. Sin embargo, el resultado no es verídico en campos arbitrarios.

10. Demostrar el corolario del Teorema 4.3.

4.3 PROPIEDADES DE LOS DETERMINANTES

Existen varias propiedades importantes que son de gran utilidad al evaluar un determinante de una matriz dada. Estas se resumen en el siguiente teorema.

Teorema 4.6. *Cualquier determinante* δ *en* $\mathsf{M}_{n\times n}(F)$ *tiene las siguientes propiedades:*

(a) *Si* B *es una matriz obtenida a partir de* A *al multiplicar cada elemento de un mismo renglón de* A *por un escalar* c, *entonces* $\delta(B) = c\cdot\delta(A)$.
(b) *Si dos renglones de* A *son idénticos, entonces* $\delta(A) = 0$.
(c) *Si* B *es una matriz obtenida a partir de* A *intercambiando dos renglones, entonces* $\delta(B) = -\delta(A)$.
(d) *Si un renglón de* A *consta totalmente de elementos nulos, entonces* $\delta(A) = 0$.
(e) *Si* B *es una matriz obtenida a partir de* A *al sumar un múltiplo del renglón* i *al renglón* $j (i \neq j)$, *entonces* $\delta(B) = \delta(A)$.

DEMOSTRACIÓN. La propiedad (a) es una consecuencia del hecho de que δ es una función n-lineal, mientras que las propiedades (b) y (c) son consecuencias del Teorema 4.4.

(d) Supóngase que A_i, el renglón i de A, consta totalmente de elementos nulos. Entonces

$$\delta(A) = \delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ 0A_i \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} = 0\cdot\delta\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} = 0.$$

(e) Sea B obtenida a partir de $A \in \mathsf{M}_{n\times n}(F)$ sumando c veces el renglón i al renglón j. Supóngase por razones de argumento que $i < j$. Por tanto, si

$$A = \begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ \cdot \\ \cdot \\ \cdot \\ A_j \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix}, \text{ entonces } \quad B = \begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ \cdot \\ \cdot \\ \cdot \\ cA_i + A_j \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix}.$$

Y así

$$\delta(B) = c \cdot \delta \begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} + \delta \begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_i \\ \cdot \\ \cdot \\ \cdot \\ A_j \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix} = c \cdot 0 + \delta(A) = \delta(A)$$

por la n-linealidad de δ y por la propiedad anterior (b). ■

Obsérvese que las propiedades (a), (c) y (e) del Teorema 4.6 muestran cómo cambia el determinante de una matriz cuando se realiza en la matriz una operación elemental en los renglones. Podemos reformular estas propiedades en términos de matrices elementales de la manera siguiente.

Corolario. *Sean* E_1, E_2 *y* E_3 *matrices elementales en* $M_{n \times n}(F)$ *respectivamente de los tipos 1, 2 y 3. Si* E_2 *se obtiene multiplicando un renglón de* I *por el escalar no nulo* c, *entonces para cualquier determinante* δ *en* $M_{n \times n}(F)$, $\delta(E_1) = -1$, $\delta(E_2) = c$, *y* $\delta E_3) = 1$.

Este corolario es uno de los ingredientes clave de la demostración de la unicidad de un determinante en $M_{n \times n}(F)$. Demostraremos ahora los dos teoremas restantes que serán necesarios para establecer la unicidad. Nuestro primer resultado calcula el determinante de cualquier matriz no invertible.

Teorema 4.7. *Sea* δ *un determinante en* $M_{n \times n}(F)$ *y sea* A *un elemento de* $M_{n \times n}(F)$ *de rango menor que* n. *Entonces* $\delta(A) = 0$.

DEMOSTRACIÓN. Como rango$(A) < n$, los renglones de A son linealmente dependientes (Corolario 2 del Teorema 3.5). Por tanto existen escalares $c_1, \ldots, c_n$, no todos cero, tales que $c_1A_1 + c_2A_2 + \ldots + c_nA_n = 0$, donde $A_1, A_2, \ldots, A_n$ son los renglones de A. Supóngase en favor del argumento que $c_1 \neq 0$; entonces

$$A_1 + c_1^{-1}c_2A_2 + \ldots + c_1^{-1}c_nA_n = 0.$$

Sea B la matriz obtenida a partir de A sumando al primer renglón el múltiplo $c_1^{-1}c_iA_i$ del renglón i para toda $i (i = 2, \ldots, n)$. Entonces el primer renglón de B consta totalmente de elementos nulos, de modo que $\delta(B) = 0$. Pero $\delta(B) = \delta(A)$ de acuerdo con la propiedad (e) del Teorema 4.6. Por tanto $\delta(A) = 0$. ■

El siguiente resultado establece el hecho final necesario para demostrar la unicidad de un determinante en $\mathsf{M}_{n \times n}(F)$, que un determinante se comporta adecuadamente respecto a la multiplicación matricial. Sin embargo, este teorema es de considerable importancia por su propio derecho. En especial su segundo corolario, que proporciona una prueba determinante para la invertibilidad de una matriz, será utilizado con frecuencia en los próximos capítulos.

Lema. *Si* E *es una matriz elemental de* $n \times n$ *con elementos de* F *y si* δ *es un determinante en* $\mathsf{M}_{n \times n}(F)$, *entonces* $\delta(EB) = \delta(E) \cdot \delta(B)$ *para cualquier* $B \in \mathsf{M}_{n \times n}(F)$.

DEMOSTRACIÓN. Supóngase que al multiplicar por la izquierda por E se intercambian dos renglones de B. Entonces $\delta(EB) = -\delta(B)$ de acuerdo con el Teorema 4.6(c). Pero $\delta(E) = -1$ por el corolario al Teorema 4.6; entonces $\delta(EB) = \delta(E) \cdot \delta(B)$. Demostraciones semejantes establecen el resultado para la multiplicación de un renglón de B por un escalar no nulo o para la suma de un múltiplo de un renglón a otro. ■

Teorema 4.8. *Sea* δ *un determinante en* $\mathsf{M}_{n \times n}(F)$, *y sean* A *y* B *elementos cualesquiera de* $\mathsf{M}_{n \times n}(F)$. *Entonces* $\delta(AB) = \delta(A) \cdot \delta(B)$.

DEMOSTRACIÓN. Si rango$(A) < n$, entonces de acuerdo con el Teorema 3.6 rango$(AB) \leq$ rango$(A) < n$. Por tanto por el Teorema 4.7 $\delta(AB) = 0$ y $\delta(A) = 0$. Y en este caso $\delta(AB) = \delta(A) \cdot \delta(B)$.

Si rango$(A) = n$, A es invertible y por tanto es el resultado del producto de matrices elementales (Corolario 3 del Teorema 3.5). Sea $A = E_m \ldots E_1$, donde cada E_i es una matriz elemental. Entonces por el lema tenemos

$$\delta(AB) = \delta(E_m \ldots E_1B) = \delta(E_m) \cdot \delta(E_{m-1} \ldots E_1B) = \ldots$$

$$= \delta(E_m) \cdot \ldots \cdot \delta(E_1) \cdot \delta(B) = \delta(E_m \ldots E_1) \cdot \delta(B) = \delta(A) \cdot \delta(B). \quad ■$$

Corolario 1. *Sea δ un determinante en* $M_{n\times n}(F)$, *y sea* $A \in M_{n\times n}(F)$ *invertible. Entonces* $\delta(A) \neq 0$ *y* $\delta(A^{-1}) = [\delta(A)]^{-1}$.

DEMOSTRACIÓN. En virtud del Teorema 4.8 se tiene que

$$\delta(A) \cdot \delta(A^{-1}) = \delta(AA^{-1}) = \delta(I_n) = 1.$$

De manera que $\delta(A) \neq 0$ y $\delta(A^{-1}) = [\delta(A)]^{-1}$. ■

Corolario 2. *Sea δ un determinante en* $M_{n\times n}(F)$ *y sea* $A \in M_{n\times n}(F)$. *Entonces las siguientes condiciones son equivalentes:*

(a) $\delta(A) = 0$.
(b) A *no es invertible.*
(c) *rango*$(A) < n$.

DEMOSTRACIÓN. El corolario anterior muestra que si $\delta(A) = 0$, entonces A no es invertible. De aquí que la condición (a) implica la condición (b).

El que la condición (b) implique la condición (c) se deriva de una observación previa en la página 146.

Finalmente, el Teorema 4.7 muestra que la condición (c) implica la condición (a). ■

Se demostró en los Teoremas 4.1 y 4.2 que existe exactamente un determinante en $M_{2\times 2}(F)$. Ahora podemos demostrar un resultado semejante para $M_{n\times n}(F)$.

Teorema 4.9. *Existe exactamente un determinante en* $M_{n\times n}(F)$.

DEMOSTRACIÓN. La existencia de un determinante en $M_{n\times n}(F)$ se demostró en el Corolario 2 del Teorema 4.5.

Completaremos la demostración, estableciendo que si δ_1 y δ_2 son determinantes en $M_{n\times n}(F)$, entonces $\delta_1 = \delta_2$. Sea A una matriz arbitraria de $n \times n$ con elementos de F. Si rango$(A) < n$ entonces $\delta_1(A) = \delta_2(A) = 0$ por el Corolario 2 del Teorema 4.8. Si rango$(A) = n$, entonces A es invertible y por tanto es el producto de matrices elementales (Corolario 3 del Teorema 3.5). Sea $A = E_m \ldots E_1$, donde cada E_i es una matriz elemental. Como $\delta_1(E_i) = \delta_2(E_i)$ para cada $i (1 \leq i \leq m)$ por el corolario al Teorema 4.6,

$$\begin{aligned}\delta_1(A) &= \delta_1(E_m \ldots E_1) = \delta_1(E_m) \ldots \delta_1(E_1)\\ &= \delta_2(E_m) \ldots \delta_2(E_1) = \delta_2(E_m \ldots E_1) = \delta_2(A)\end{aligned}$$

por el Teorema 4.8. Por tanto $\delta_1 = \delta_2$. ■

De aquí en adelante denotaremos al único determinante en $M_{n\times n}(F)$ por *det.*

Corolario. *Sea* $A \in M_{n\times n}(F)$. *Para toda* $j(1 \leq j \leq n)$

$$det(A) = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \cdot det(\tilde{A}_{ij}),$$

donde $\tilde{A}_{ij}$ *es la matriz de* $(n-1) \times (n-1)$ *obtenida de* A *al suprimir el renglón* i *y la columna* j.

Así, el determinante de una matriz de $n \times n$ puede evaluarse por expansión sobre cualquier columna; si $n > 2$ la expansión resultante contendrá n determinantes de matrices de $(n-1) \times (n-1)$. El determinante de cada una de esas matrices de $(n-1) \times (n-1)$ puede expandirse sobre cualquier columna y este proceso puede continuar hasta que una expansión implique únicamente determinantes de matrices de 2×2, que pueden ser evaluados mediante la expresión $\det(A) = A_{11}A_{22} - A_{12}A_{21}$.

Obsérvese sin embargo que la evaluación de $\det(\tilde{A}_{ij})$ puede evitarse siempre que $A_{ij} = 0$, pues el producto $A_{ij} \cdot \det(\tilde{A}_{ij})$ será cero independientemente del valor del determinante. Por tanto, es benéfico expandir sobre una columna que tenga tantos ceros como sea posible. Ilustraremos este procedimiento con dos ejemplos.

Ejemplo 9. Sea A el siguiente elemento de $M_{4\times 4}(F)$:

$$\begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

Para minimizar el cálculo requerido para evaluar $\det(A)$, expandiremos sobre la segunda columna. Entonces

$$\begin{aligned} \det(A) &= \sum_{i=1}^{4} (-1)^{i+2} A_{i2} \cdot \det(\tilde{A}_{i2}) \\ &= (-1)^{1+2} \cdot 1 \cdot \det\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} + (-1)^{2+2} \cdot 0 \cdot \det\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \\ &\quad + (-1)^{3+2} \cdot 0 \cdot \det\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} + (-1)^{4+2} \cdot 1 \cdot \det\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \\ &= (-1) \cdot 1 \cdot 0 + 0 + 0 + 1 \cdot 1 \cdot \det\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \det\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \end{aligned}$$

(La primera de las cuatro matrices de 3×3 tiene dos renglones idénticos, por lo que su determinante es cero.) Evaluaremos ahora el determinante que queda expandiendo sobre la primera columna. Entonces

$$\begin{aligned}\det(A) &= \det\begin{pmatrix}1 & 0 & 1\\ 1 & 1 & 1\\ 0 & 1 & 1\end{pmatrix}\\ &= (-1)^{1+1}\cdot 1\cdot\det\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix} + (-1)^{2+1}\cdot 1\cdot\det\begin{pmatrix}0 & 1\\ 1 & 1\end{pmatrix}\\ &\quad + (-1)^{3+1}\cdot 0\cdot\det\begin{pmatrix}0 & 1\\ 1 & 1\end{pmatrix}\\ &= 1\cdot 1\cdot 0 + (-1)\cdot 1\cdot(-1) + 0 = 1.\end{aligned}$$

Ahora sea B el siguiente elemento de $\mathsf{M}_{5\times 5}(R)$:

$$\begin{pmatrix}1 & -1 & 0 & 0 & 0\\ -2 & 0 & -3 & 1 & 6\\ 5 & -4 & 0 & 2 & 0\\ 0 & 3 & 0 & -1 & 4\\ -9 & 8 & 0 & 0 & 0\end{pmatrix}.$$

Expandiendo sucesivamente sobre la tercera, cuarta y tercera columnas vemos que

$$\begin{aligned}\det(B) &= (-1)^{2+3}\cdot(-3)\cdot\det\begin{pmatrix}1 & -1 & 0 & 0\\ 5 & -4 & 2 & 0\\ 0 & 3 & -1 & 4\\ -9 & 8 & 0 & 0\end{pmatrix}\\ &= 3(-1)^{3+4}\cdot 4\cdot\det\begin{pmatrix}1 & -1 & 0\\ 5 & -4 & 2\\ -9 & 8 & 0\end{pmatrix}\\ &= -12\cdot(-1)^{2+3}\cdot 2\cdot\det\begin{pmatrix}1 & -1\\ -9 & 8\end{pmatrix}\\ &= 24[1\cdot 8 - (-1)(-9)] = 24(-1) = -24.\end{aligned}$$

Como estos ejemplos lo sugieren, el proceso de cálculo de un determinante es a menudo tedioso aun cuando se encuentren presentes elementos nulos; sin elementos nulos la evaluación de un determinante por expansión sobre una columna es muy ineficiente. En vez de este procedimiento podemos utilizar la propiedad (e) del Teorema 4.6 para cambiar la matriz A en una matriz B que tenga el mismo determinante que A y que tenga elementos nulos en una o más columnas. Este es en esencia el

mismo proceso que se utilizó para reducir A a la forma escalonada. A continuación se tienen ejemplos de esta técnica.

Ejemplo 10. Sea A el siguiente elemento de $\mathsf{M}_{4\times 4}(R)$:

$$\begin{pmatrix} 1 & -1 & 2 & -1 \\ -3 & 4 & 1 & -1 \\ 2 & -5 & -3 & 8 \\ -2 & 6 & -4 & 1 \end{pmatrix}.$$

Entonces

$$\det(A) = \det\begin{pmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{pmatrix} = \det\begin{pmatrix} A_1 \\ 3A_1 + A_2 \\ -2A_1 + A_3 \\ 2A_1 + A_4 \end{pmatrix}$$

$$= \det\begin{pmatrix} 1 & -1 & 2 & -1 \\ 0 & 1 & 7 & -4 \\ 0 & -3 & -7 & 10 \\ 0 & 4 & 0 & -1 \end{pmatrix} = (-1)^{1+1}\cdot 1\cdot\det\begin{pmatrix} 1 & 7 & -4 \\ -3 & -7 & 10 \\ 4 & 0 & -1 \end{pmatrix}$$

$$= \det\begin{pmatrix} 1 & 7 & -4 \\ -2 & 0 & 6 \\ 4 & 0 & -1 \end{pmatrix} = (-1)^{1+2}\cdot 7\cdot\det\begin{pmatrix} -2 & 6 \\ 4 & -1 \end{pmatrix}$$

$$= -7[(-2)(-1) - 6\cdot 4] = -7(-22) = 154.$$

Ejemplo 11. Sea A el siguiente elemento de $\mathsf{M}_{4\times 4}(R)$:

$$\begin{pmatrix} 1 & -1 & 2 & 1 \\ 2 & -1 & -1 & 4 \\ -4 & 5 & -10 & -6 \\ 3 & -2 & 10 & -1 \end{pmatrix}.$$

Introduciremos elementos nulos mediante el uso del Teorema 4.6(e) de tal modo que A se transforme en una matriz triangular superior que tenga el mismo determinante que A. El determinante de la matriz triangular superior se calculará entonces por expansiones sucesivas sobre la primera columna.

$$\det(A) = \det\begin{pmatrix} 1 & -1 & 2 & 1 \\ 2 & -1 & -1 & 4 \\ -4 & 5 & -10 & -6 \\ 3 & -2 & 10 & -1 \end{pmatrix} = \det\begin{pmatrix} 1 & -1 & 2 & 1 \\ 0 & 1 & -5 & 2 \\ 0 & 1 & -2 & -2 \\ 0 & 1 & 4 & -4 \end{pmatrix}$$

$$= \det\begin{pmatrix} 1 & -1 & 2 & 1 \\ 0 & 1 & -5 & 2 \\ 0 & 0 & 3 & -4 \\ 0 & 0 & 9 & -6 \end{pmatrix} = \det\begin{pmatrix} 1 & -1 & 2 & 1 \\ 0 & 1 & -5 & 2 \\ 0 & 0 & 3 & -4 \\ 0 & 0 & 0 & 6 \end{pmatrix}$$

$$= 1 \cdot \det\begin{pmatrix} 1 & -5 & 2 \\ 0 & 3 & -4 \\ 0 & 0 & 6 \end{pmatrix} = 1 \cdot 1 \cdot \det\begin{pmatrix} 3 & -4 \\ 0 & 6 \end{pmatrix}$$

$$= 1 \cdot 1 \cdot 3 \cdot 6 = 18.$$

Hasta ahora, el papel jugado por los renglones y columnas de una matriz en el estudio de los determinantes ha sido muy diferente —un determinante se definió como una función en $\mathsf{M}_{n\times n}(F)$ que satisface ciertas propiedades que involucran a los renglones de una matriz, mientras que la evaluación del determinante se lleva a cabo mediante la expansión sobre las columnas de la matriz. Estos papeles son reversibles y ahora verificaremos este hecho demostrando que los determinantes de A y A^t son iguales. (Como los renglones de A son las columnas de A^t y viceversa, este resultado será suficiente para demostrar que los papeles desempeñados por los renglones y las columnas son intercambiables.)

Teorema 4.10. *Para cualquier matriz* A *de* $n \times n$ $\det(A^t) = \det(A)$.

DEMOSTRACIÓN. Si A no es invertible, entonces rango$(A) < n$. Pero como rango(A^t) = rango(A) (Corolario 2 del Teorema 3.5) se tiene que A^t no es invertible, y entonces, en este caso, $\det(A) = 0 = \det(A^t)$.

Si A es invertible, entonces $A = E_m \ldots E_1$, donde $E_1, \ldots, E_m$ son matrices elementales. Como $\det(E_i^t) = \det(E_i)$ para cada i (véase el Ejercicio 5), tenemos que

$$\begin{aligned}\det(A^t) &= \det(E_1^t \ldots E_m^t) = \det(E_1^t) \ldots \det(E_m^t) \\ &= \det(E_1) \ldots \det(E_m) = \det(E_m) \ldots \det(E_1) \\ &= \det(E_m \ldots E_1) = \det(A). \quad \blacksquare\end{aligned}$$

Corolario. *Cualquier argumento sobre determinantes que involucre a los renglones de una matriz puede ser enunciado de nuevo en términos de las columnas de la matriz, y cualquier argumento que involucre a las columnas de una matriz puede ser enunciado de nuevo en términos de los renglones de la matriz. En particular, si* A *es una matriz de* $n \times n$,

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \cdot \det(\tilde{A}_{ij}),$$

donde $\tilde{A}_{ij}$ *es la matriz de* $(n-1) \times (n-1)$ *obtenida a partir de* A *al suprimir el renglón* i *y la columna* j.

Ejemplo 12. Sea A el siguiente elemento de $M_{4\times4}(R)$:

$$\begin{pmatrix} 4 & -1 & -3 & 6 \\ -2 & 3 & 1 & 4 \\ 0 & 5 & 0 & 0 \\ 1 & 2 & 3 & -1 \end{pmatrix}.$$

En este caso el cálculo requerido para evaluar $\det(A)$ puede ser minimizado al expandir sobre el tercer renglón, y entonces

$$\det(A) = -5\det\begin{pmatrix} 4 & -3 & 6 \\ -2 & 1 & 4 \\ 1 & 3 & -1 \end{pmatrix} = -5\det\begin{pmatrix} 0 & -15 & 10 \\ 0 & 7 & 2 \\ 1 & 3 & -1 \end{pmatrix}$$

$$= -5\det\begin{pmatrix} -15 & 10 \\ 7 & 2 \end{pmatrix} = -5[(-15)\cdot 2 - 10\cdot 7] = 500.$$

Nuestro resultado final nos permite evaluar fácilmente el determinante de una matriz triangular. Este resultado hace de la técnica utilizada en el Ejemplo 11 un método muy eficiente para evaluar determinantes.

Teorema 4.11. *Si* A *es una matriz triangular de* $n \times n$, *entonces* $det(A) = A_{11}A_{22} \ldots A_{nn}$; *esto es, el determinante de* A *es el producto de los elementos de* A *ubicados en la diagonal.*

DEMOSTRACIÓN. Sea A una matriz triangular superior de $n \times n$. La demostración se hace por inducción sobre n. Si $n = 2$, entonces A tiene la forma

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix},$$

y entonces $\det(A) = A_{11}A_{22} - A_{12}\cdot 0 = A_{11}A_{22}$, lo que demuestra el teorema para matrices triangulares superiores si $n = 2$.

Supóngase que el teorema es cierto para matrices triangulares superiores de $(n-1) \times (n-1)$ y sea A una matriz triangular superior de $n \times n$. Entonces A tiene la forma

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1(n-1)} & A_{1n} \\ 0 & A_{22} & \cdots & A_{2(n-1)} & A_{2n} \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & A_{nn} \end{pmatrix}.$$

Al expandir sobre la primera columna, vemos que

$$\det(A) = A_{11}\cdot\det\begin{pmatrix} A_{22} & \cdots & A_{2(n-1)} & A_{2n} \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ 0 & \cdots & 0 & A_{nn} \end{pmatrix}$$

$$= A_{11}\cdot(A_{22}\cdots A_{nn})$$

por la hipótesis de inducción. Esto completa la inducción y demuestra el teorema para matrices triangulares superiores.

Si A es una matriz triangular inferior, entonces A^t es una matriz triangular superior. Por tanto, la primera parte de esta demostración y el Teorema 4.10 implican que

$$\det(A) = \det(A^t) = (A^t)_{11} \ldots (A^t)_{nn} = A_{11} \ldots A_{nn}. \quad \blacksquare$$

Tal como en la Sección 4.1, es posible interpretar geométricamente el determinante de un elemento A en $\mathsf{M}_{n\times n}(R)$. Si $A_1, \ldots, A_n$ son los n renglones de A, podemos interpretar

$$\left|\det\begin{pmatrix} A_1 \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{pmatrix}\right|$$

como el volumen n-dimensional (la generalización del área en R^2 y del volumen en R^3) del paralelepípedo que tiene a los vectores $A_1, \ldots, A_n$ como aristas adyacentes. (Para una demostración de este resultado véase a Serge Lang, *Análisis I*, Addison-Wesley, 1968, pp. 413-418.)

En nuestra anterior exposición del significado geométrico del determinante formado a partir de los vectores en una base ordenada para R^2, vimos también que este determinante es positivo si y sólo si la base induce un sistema coordenado derecho. Una aseveración similar es verídica en R^n. Específicamente, si γ es cualquier base ordenada para R^n y β es la ordenada estándar para R^n, entonces γ induce un sistema coordenado derecho si y sólo si $\det(Q) > 0$, donde Q es la matriz de cambio de coordenadas que permite pasar de coordenadas de γ a coordenadas de β. Entonces, por ejemplo,

$$\gamma = \left\{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right\}$$

induce un sistema coordinado izquierdo en R^3 puesto que

$$\det\begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = -2 < 0,$$

mientras que

$$\gamma' = \left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

induce un sistema coordenado derecho en R^3 porque

$$\det\begin{pmatrix} 1 & -2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = 5 > 0.$$

Más generalmente, si β y γ son dos bases ordenadas cualesquiera de R^n, entonces los sistemas coordenados inducidos por β y γ tienen la misma orientación (ambos derechos o ambos izquierdos) si y sólo si $\det(Q) > 0$, donde Q es la matriz de cambio de coordenadas que permite el paso de las coordenadas de γ a las de β.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

(a) Si dos renglones de A son idénticos, entonces $\det(A) = 0$.

(b) Si B es una matriz obtenida a partir de A mediante el intercambio de dos renglones, entonces $\det(B) = -\det(A)$.

(c) Si B es una matriz obtenida a partir de A multiplicando un renglón de A por un escalar c, entonces $\det(A) = \det(B)$.

(d) Si B es una matriz obtenida a partir de A sumando un múltiplo escalar del renglón i al renglón $j (i \neq j)$, entonces $\det(A) = \det(B)$.

(e) Si E es una matriz elemental, entonces $\det(E) = \pm 1$.

(f) Si $A, B \in \mathsf{M}_{n\times n}(F)$, entonces $\det(AB) = \det(A) \cdot \det(B)$.

(g) Una matriz M es invertible si y sólo si $\det(M) = 0$.

(h) Una matriz $M \in \mathsf{M}_{n\times n}(F)$ tiene rango n si y sólo si $\det(M) \neq 0$.

(i) El determinante de una matriz puede ser evaluado por expansión sobre cualquier renglón o columna.

(j) $\det(A^t) = -\det(A)$.

(k) El determinante de una matriz diagonal es igual al producto de los elementos de la diagonal.

2. Evaluar cada uno de los siguientes determinantes por el método que se indica.

(a) Por expansión sobre la segunda columna

$$\begin{pmatrix} 2 & -1 & 4 \\ 3 & 6 & 1 \\ -1 & 2 & 3 \end{pmatrix}$$

(b) Por expansión sobre el primer renglón

$$\begin{pmatrix} -3 & 1 & 2 \\ 1 & -1 & 4 \\ 5 & 6 & 1 \end{pmatrix}$$

(c) Por expansión sobre la tercera columna

$$\begin{pmatrix} 1 & 2 & 1 & 1 \\ -1 & 1 & 0 & 1 \\ 2 & 0 & 0 & -1 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

(d) Por expansión sobre el cuarto renglón

$$\begin{pmatrix} 1 & 2 & 0 & -1 \\ 1 & -1 & 0 & 2 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 \end{pmatrix}$$

3. Evaluar los determinantes de las matrices siguientes por cualquier método permitido. En cada caso C es el campo de escalares.

(a) $\begin{pmatrix} 4 & -7 & 3 \\ 1 & 2 & -1 \\ -3 & 4 & 5 \end{pmatrix}$ (b) $\begin{pmatrix} 9 & 0 & 0 \\ 4 & 8 & 0 \\ 3 & 2 & 7 \end{pmatrix}$ (c) $\begin{pmatrix} 4 & -5 & 2 \\ 2 & 8 & 1 \\ 6 & -1 & 3 \end{pmatrix}$

(d) $\begin{pmatrix} -2+i & -1 & 5i \\ 3 & 3+2i & -2i \\ 4i & 0 & 1+i \end{pmatrix}$ (e) $\begin{pmatrix} 1 & 2 & -1 & -1 \\ -3 & 0 & 2 & 1 \\ 2 & -1 & 5 & 4 \\ -1 & 6 & 3 & 3 \end{pmatrix}$

(f) $\begin{pmatrix} 2 & 0 & -1 & 3 \\ -4 & 3 & 5 & 1 \\ 1 & 6 & 0 & 2 \\ 0 & -5 & 3 & 7 \end{pmatrix}$ (g) $\begin{pmatrix} -1+3i & 2i & 6 & 0 \\ 4 & 0 & 3+i & 4i \\ 0 & 1-2i & 0 & 2-i \\ 2i & 5 & 0 & 1+i \end{pmatrix}$

4. Demostrar que una matriz triangular de $n \times n$ es invertible si y sólo si en la diagonal no se encuentra ningún cero.

5. Completar la demostración del Teorema 4.10 demostrando que si E es una matriz elemental, entonces $\det(E^t) = \det(E)$. *Sugerencia:* E^t es una matriz elemental del mismo tipo que E.

6. Demostrar que si $A \in \mathsf{M}_{n \times n}(F)$, entonces $\det(cA) = c^n \det(A)$ para cualquier escalar c.

7. (a) Una matriz B en $M_{n\times n}(R)$ se llama *ortogonal* si $BB^t = I$. Demostrar que si B es ortogonal, entonces $\det(B) = \pm 1$.
   (b) Una matriz B en $M_{n\times n}(C)$ se llama *unitaria* si $BB^* = I$ donde $(B^*)_{ij} = \overline{B_{ji}}$, el complejo conjugado de B_{ji}. Demostrar que si B es unitaria entonces $|\det(B)| = 1$. *Sugerencia:* Demostrar primero que $\det(\overline{B}) = \overline{\det(B)}$.

8. Una matriz B en $M_{n\times n}(C)$ se llama *antisimétrica* si $B^t = -B$. Demostrar que si $B \in M_{n\times n}(C)$ es antisimétrica y n es impar, entonces $\det(B) = 0$.

9.† Sea $A \in M_{n\times n}(F)$, y sea m tal que $1 \leq m \leq n$. Sean

$$B_1 = \begin{pmatrix} A_{11} & \cdots & A_{1m} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ A_{m1} & \cdots & A_{mm} \end{pmatrix}, \qquad B_2 = \begin{pmatrix} A_{1(m+1)} & \cdots & A_{1n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ A_{m(m+1)} & \cdots & A_{mn} \end{pmatrix},$$

$$B_3 = \begin{pmatrix} A_{(m+1)(m+1)} & \cdots & A_{(m+1)n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ A_{n(m+1)} & \cdots & A_{nn} \end{pmatrix},$$

y O sea la matriz nula de $(n-m) \times m$. A se puede escribir simbólicamente como

$$A = \begin{pmatrix} B_1 & B_2 \\ O & B_3 \end{pmatrix}.$$

Demostrar que $\det(A) = \det(B_1) \cdot \det(B_3)$.

10. Sea $\beta = \{x_1, \ldots, x_n\}$ un subconjunto de F^n que contiene n vectores diferentes, y sea B el elemento de $M_{n\times n}(F)$ cuya j-ésima columna es el vector x_j. Demostrar que β es una base para F^n si y sólo si $\det(B) \neq 0$.

11. Complete la demostración del lema del Teorema 4.8.

12. Recuérdese la transformación lineal $T\colon P_n(F) \to F^{n+1}$ definida en el Ejercicio 20 de la Sección 2.4 por $T(f) = (f(c_0), \ldots, f(c_n))$, donde $c_0, \ldots, c_n$ son elementos distintos de un campo infinito F. Sea $\beta = \{1, x, x^2, \ldots, x^n\}$ una base ordenada de $P_n(F)$ y γ la base ordenada normal para F^{n+1}.
   (a) Calcular $M = [T]_\beta^\gamma$. Una matriz que tiene la forma de M se llama *matriz de Vandermonde*.
   (b) Demostrar que $\det(M) \neq 0$ utilizando el Ejercicio 20 de la Sección 2.4.

(c) Demostrar que

$$\det(M) = \prod_{0 \leq i < j \leq n} (c_j - c_i),$$

el producto de todos los términos de la forma $c_j - c_i$ para $0 \leq i < j \leq n$.

13. Sea $A \in M_{m \times n}(F)$ no nula. Para toda $m(1 \leq m \leq n)$ una submatriz de $m \times m$ se obtiene de A suprimiendo $n - m$ renglones y $n - m$ columnas cualesquiera de A. Sea $k(1 \leq k \leq n)$ el mayor entero tal que alguna submatriz de $k \times k$ de A tiene un determinante no nulo. Demostrar que $\text{rango}(A) = k$.

14. Utilizar los resultados de esta sección para demostrar el Ejercicio 8 de la Sección 2.4: Si A y B son matrices de $m \times n$ tales que $AB = I_n$, entonces A es invertible (y por tanto $B = A^{-1}$).

15. Demostrar que si A y B son matrices similares, entonces $\det(A) = \det(B)$.

4.4 LA ADJUNTA CLASICA Y LA REGLA DE CRAMER

En esta sección definiremos la adjunta clásica de una matriz de $n \times n$ y la utilizaremos para calcular la inversa de una matriz. También obtendremos la regla de Cramer, la cual permite utilizar determinantes para resolver un sistema de ecuaciones lineales cuando éste tenga una matriz de coeficientes invertible. Nuestra herramienta principal será el siguiente teorema que muestra las consecuencias de una expansión de una matriz por elementos de una columna y cofactores.

Teorema 4.12. *Sea* A *una matriz de* $n \times n$ *y sea* c_{ij} *el cofactor de* $A_{ij}(1 \leq i, j \leq n)$. *Entonces*

$$\sum_{i=1}^{n} A_{ij} \cdot c_{ik} = \delta_{jk} \cdot \det(A),$$

donde δ_{jk} *es la delta de Kronecker.*

DEMOSTRACIÓN. Si $j = k$, la ecuación se obtiene a partir del corolario del Teorema 4.9. Supóngase que $j \neq k$ y sea B la matriz que tiene todas sus columnas idénticas a las columnas correspondientes de A excepto la columna k-ésima, B^k, que es idéntica a la columna j de A. Entonces $B^k = B^j = A^j$, y el cofactor de B_{ik} es c_{ik}. Ahora bien, $\det(B) = 0$ puesto que dos columnas de B son idénticas; pero, expandiendo B sobre la columna k, tenemos también

$$\det(B) = \sum_{i=1}^{n} B_{ik}c_{ik} = \sum_{i=1}^{n} A_{ij}c_{ik}.$$

Y por tanto

$$\sum_{i=1}^{n} A_{ij}c_{ik} = 0 \qquad \text{si } j \neq k. \quad \blacksquare$$

Corolario 1. *Si* A *y* c_{ij} *son como en el enunciado del Teorema 4.12, entonces*

$$\sum_{j=1}^{n} A_{ij}c_{kj} = \delta_{ik} \cdot det(A).$$

DEMOSTRACIÓN. Ejercicio.

Definición. *Sea* A *una matriz de* n × n. *La matriz de* n × n adj(A) *cuyo elemento del renglón* i *y columna* j *es el cofactor de* A_{ji} *se llama la* adjunta clásica *de* A. *(Entonces* adj A = C^t, *donde* C_{ij} *es el cofactor de* A_{ij}.)

Ejemplo 13. Sean A y B los siguientes elementos de $M_{3\times 3}(R)$:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & -1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 3 & -1 \\ -1 & -4 & 2 \\ 2 & 0 & 1 \end{pmatrix}.$$

Entonces

$$\text{adj } A = \begin{pmatrix} -1 & 2 & 1 \\ 5 & -4 & 1 \\ 2 & 2 & -2 \end{pmatrix}^t = \begin{pmatrix} -1 & 5 & 2 \\ 2 & -4 & 2 \\ 1 & 1 & -2 \end{pmatrix},$$

y

$$\text{adj } B = \begin{pmatrix} -4 & 5 & 8 \\ -3 & 3 & 6 \\ 2 & -1 & -1 \end{pmatrix}^t = \begin{pmatrix} -4 & -3 & 2 \\ 5 & 3 & -1 \\ 8 & 6 & -1 \end{pmatrix}.$$

Corolario 2. *Para toda matriz* A *de* n × n (*adj* A)A = [*det*(A)]I.

DEMOSTRACIÓN. Sea $A \in M_{n\times n}(F)$, y sea c_{ij} el cofactor de A_{ij}. Entonces $(\text{adj } A)_{ji} = c_{ij}$, y por tanto, el elemento de $(\text{adj } A)A$ para el j-ésimo renglón y k-ésima columna es

$$\sum_{i=1}^{n} (\text{adj } A)_{ji}A_{ik} = \sum_{i=1}^{n} c_{ij}A_{ik} = \delta_{jk} \cdot \det(A)$$

por el Corolario 1. Así $(\text{adj } A)A = [\det(A)]I$. ■

Ejemplo 14. Sea A tal como en el Ejemplo 13. Al expandir A sobre el segundo renglón, vemos que

$$\det(A) = (-1)\cdot 1 \cdot [2(-1) - 3\cdot 1] + (-1)\cdot 1 \cdot (1\cdot 1 - 2\cdot 1) = 6.$$

Y

$$(\text{adj } A)A = \begin{pmatrix} -1 & 5 & 2 \\ 2 & -4 & 2 \\ 1 & 1 & -2 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & -1 \end{pmatrix} = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{pmatrix}$$
$$= [\det(A)]I.$$

De una manera análoga, si B es como en el Ejemplo 13, entonces al expandir B sobre el tercer renglón vemos que

$$\det(B) = 2[3 \cdot 2 - (-1)(-4)] + 1[1(-4) - 3(-1)] = 3.$$

Más aún,

$$(\text{adj } B)B = \begin{pmatrix} -4 & -3 & 2 \\ 5 & 3 & -1 \\ 8 & 6 & -1 \end{pmatrix}\begin{pmatrix} 1 & 3 & -1 \\ -1 & -4 & 2 \\ 2 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$
$$= [\det(B)]I.$$

Corolario 3. *Si* A *es una matriz invertible, entonces*

$$A^{-1} = [\det(A)]^{-1}(\text{adj } A).$$

DEMOSTRACIÓN. Si A es invertible, entonces $\det(A) \neq 0$ (Corolario 1 del Teorema 4.8). Ahora bien, $(\text{adj } A)A = [\det(A)]I$ por el Corolario 2 y por tanto $[\det(A)]^{-1}(\text{adj } A)A = I$. Y entonces $[\det(A)]^{-1}(\text{adj } A) = A^{-1}$. ■

Ejemplo 15. Continuando con el Ejemplo 14, tenemos

$$A^{-1} = [\det(A)]^{-1}(\text{adj } A) = \frac{1}{6}\begin{pmatrix} -1 & 5 & 2 \\ 2 & -4 & 2 \\ 1 & 1 & -2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{6} & \frac{5}{6} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{6} & \frac{1}{6} & -\frac{1}{3} \end{pmatrix}$$

y

$$B^{-1} = [\det(B)]^{-1}(\text{adj } B) = \frac{1}{3}\begin{pmatrix} -4 & -3 & 2 \\ 5 & 3 & -1 \\ 8 & 6 & -1 \end{pmatrix} = \begin{pmatrix} -\frac{4}{3} & -1 & \frac{2}{3} \\ \frac{5}{3} & 1 & -\frac{1}{3} \\ \frac{8}{3} & 2 & -\frac{1}{3} \end{pmatrix}.$$

Concluiremos esta sección con una exposición de la regla de Cramer, que proporciona un método interesante para resolver ecuaciones matriciales de la forma $AX = B$, donde A es una matriz invertible. Este método, sin embargo, es extremadamente ineficiente, puesto que si A es una matriz de $n \times n$, la solución del sistema $AX = B$ por medio de la regla de Cramer requiere de la evaluación de $n + 1$ determinantes de matrices de $n \times n$. (Por comparación, el método de solución presentado en la Sección 3.4 es un modo más eficiente para resolver tales sistemas. Por lo tanto, la regla de Cramer es más bien de interés teórico y estético, que práctico.)

Teorema 4.13. *(Regla de Cramer.) Sea* AX = B *la ecuación matricial de un sistema de* n *ecuaciones lineales con* n *incógnitas, donde* $X = (x_1, \dots, x_n)^t$ *y* $B = (b_1, \dots, b_n)^t$. *Si* $det(A) \neq 0$ *el sistema tiene una solución única y para toda* $i(1 \leq i \leq n)$

$$x_i = [det(A)]^{-1} \cdot det(M_i),$$

donde M_i *es la matriz de* $n \times n$ *obtenida a partir de* A *reemplazando la* i-*ésima columna de* A *por* B.

DEMOSTRACIÓN. Por el Corolario 2 del Teorema 4.8, $\det(A) \neq 0$ implica que A es invertible. Por tanto, de acuerdo con el Teorema 3.9, la ecuación matricial $AX = B$ tiene una solución única. Multiplicando esta ecuación por la izquierda por adj A y utilizando el Corolario 2 del Teorema 4.12 se tiene

$$[\det(A)]IX = (\text{adj } A)AX = (\text{adj } A)B.$$

Examinando las coordenadas i-ésimas de los vectores columna $[\det(A)]X = (\text{adj } A)B$, puede verse que

$$[\det(A)]x_i = \sum_{j=1}^{n} (\text{adj } A)_{ij} b_j = \sum_{j=1}^{n} c_{ji} b_j,$$

donde c_{ji} es el cofactor de A_{ji}. Pero

$$\sum_{j=1}^{n} c_{ji} b_j$$

es la expansión de M_i sobre la i-ésima columna; entonces

$$[\det(A)]x_i = \sum_{j=1}^{n} c_{ji} b_j = \det(M_i),$$

y consecuentemente

$$x_i = [\det(A)]^{-1} \cdot \det(M_i) \qquad \text{para } 1 \leq i \leq n. \qquad \blacksquare$$

Ejemplo 16. Utilizaremos la regla de Cramer para resolver la ecuación matricial $AX = B$, donde

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 1 & 1 & -1 \end{pmatrix} \quad \text{y} \quad B = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}.$$

Primero, como se vio en el Ejemplo 14, $\det(A) = 6$, por lo que se aplica la regla de Cramer. Si M_i es la matriz obtenida a partir de A substituyendo la i-ésima columna de A por B, se tiene

$$x_1 = \frac{\det(M_1)}{\det(A)} = \frac{\det\begin{pmatrix} 2 & 2 & 3 \\ 3 & 0 & 1 \\ 1 & 1 & -1 \end{pmatrix}}{\det(A)} = \frac{15}{6} = \frac{5}{2},$$

$$x_2 = \frac{\det(M_2)}{\det(A)} = \frac{\det\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 1 \\ 1 & 1 & -1 \end{pmatrix}}{\det(A)} = \frac{-6}{6} = -1,$$

y

$$x_3 = \frac{\det(M_3)}{\det(A)} = \frac{\det\begin{pmatrix} 1 & 2 & 2 \\ 1 & 0 & 3 \\ 1 & 1 & 1 \end{pmatrix}}{\det(A)} = \frac{3}{6} = \frac{1}{2}.$$

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

(a) Si una matriz se expande sobre los elementos de una columna y los cofactores diferentes de una columna el, resultado es el determinante de la matriz.

(b) Si $A \in \mathsf{M}_{n\times n}(F)$, entonces $(\operatorname{adj} A)A = I$.

(c) Todo sistema de n ecuaciones lineales con n incógnitas puede ser resuelto mediante la regla de Cramer.

(d) Sea $AX = B$ la forma matricial de un sistema de n ecuaciones lineales con n incógnitas, donde $X = (x_1, \ldots, x_n)^t$. Si $\det(A) \neq 0$ y si M_i es la matriz obtenida a partir de A al reemplazar el renglón i de A por B^t, entonces

$$x_i = [\det(A)]^{-1} \cdot \det(M_i).$$

2. Encontrar la clásica adjunta de las siguientes matrices.

(a) $\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$

(b) $\begin{pmatrix} -4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5 \end{pmatrix}$

(c) $\begin{pmatrix} 1-i & 0 & 0 \\ 4 & 3i & 0 \\ 2i & 1+4i & -1 \end{pmatrix}$

(d) $\begin{pmatrix} 3 & 6 & 7 \\ 0 & 4 & 8 \\ 0 & 0 & 5 \end{pmatrix}$

(e) $\begin{pmatrix} 7 & 1 & 4 \\ 6 & -3 & 0 \\ -3 & 5 & -2 \end{pmatrix}$ (f) $\begin{pmatrix} 3 & 2+i & 0 \\ -1+i & 0 & i \\ 0 & 1 & 3-2i \end{pmatrix}$

(g) $\begin{pmatrix} -1 & 2 & 5 \\ 8 & 0 & -3 \\ 4 & 6 & 1 \end{pmatrix}$ (h) $\begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$

3. Resolver los siguientes sistemas por medio de la regla de Cramer.

(a) $\begin{cases} a_{11}x_1 + a_{12}x_2 = b_1 \\ a_{21}x_1 + a_{22}x_2 = b_2, \end{cases}$

donde $a_{11}a_{22} - a_{12}a_{21} \neq 0$.

(b) $\begin{cases} 2x_1 + x_2 - 3x_3 = 5 \\ x_1 - 2x_2 + x_3 = 10 \\ 3x_1 + 4x_2 - 2x_3 = 0 \end{cases}$

(c) $\begin{cases} 2x_1 + x_2 - 3x_3 = 1 \\ x_1 - 2x_2 + x_3 = 0 \\ 3x_1 + 4x_2 - 2x_3 = -5 \end{cases}$

(d) $\begin{cases} x_1 - x_2 + 4x_3 = -4 \\ -8x_1 + 3x_2 + x_3 = 8 \\ 2x_1 - x_2 + x_3 = 0 \end{cases}$

(e) $\begin{cases} x_1 - x_2 + 4x_3 = -2 \\ -8x_1 + 3x_2 + x_3 = 0 \\ 2x_1 - x_2 + x_3 = 6 \end{cases}$

(f) $\begin{cases} 3x_1 + x_2 + x_3 = 4 \\ -2x_1 - x_2 = 12 \\ x_1 + 2x_2 + x_3 = -8 \end{cases}$

4. Demostrar que para cualquier $A \in \mathsf{M}_{n\times n}(F)$, $\det(\text{adj } A) = [\det(A)]^{n-1}$.

5. Sea A una matriz triangular superior invertible de $n \times n$. Demostrar que adj A es triangular superior y por tanto A^{-1} es triangular superior. Demostrar que resultados semejantes son ciertos si A es triangular inferior.

6. Demostrar el Corolario 1 del Teorema 4.12.

7. Demostrar que adj $A^t = (\text{adj } A)^t$.

4.5 RESUMEN —CONCEPTOS IMPORTANTES SOBRE DETERMINANTES

En esta sección resumiremos las propiedades importantes de los determinantes que nos serán necesarios para el resto del texto. Los resultados contenidos en esta sección fueron obtenidos en las Secciones 4.2 y 4.3; por tanto los conceptos presentados aquí serán enunciados sin demostración.

El *determinante* de una matriz A de $n \times n$ con elementos de un campo F es un elemento de F, expresado como $\det(A)$, que puede ser calculado de la siguiente manera:

1. Si A es de 1×1, entonces $\det(A) = A_{11}$, el único elemento de A.

2. Si A es de 2×2, entonces $\det(A) = A_{11}A_{22} - A_{12}A_{21}$. Así, por ejemplo,

$$\det\begin{pmatrix} -1 & 2 \\ 5 & 3 \end{pmatrix} = (-1)(3) - (2)(5) = -13.$$

3. Si A es de $n \times n$ para $n > 2$, entonces el determinante de A puede ser expresado como la suma de los productos de cada elemento de algún renglón o columna de A multiplicado por ± 1 veces el determinante de una matriz de $(n-1) \times (n-1)$ obtenida al eliminar de A el renglón y la columna que contienen el elemento en cuestión. La fórmula exacta es

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \cdot \det(\tilde{A}_{ij})$$

(si el determinante es evaluado a partir de los elementos del renglón i de A) o bien

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \cdot \det(\tilde{A}_{ij})$$

(si el determinante es evaluado a partir de los elementos de la columna j de A), donde $\tilde{A}_{ij}$ es la matriz de $(n-1) \times (n-1)$ obtenida a suprimir el renglón i y la columna j de A.

En las expresiones anteriores el escalar $(-1)^{i+j}\det(\tilde{A}_{ij})$ se llama el *cofactor* del elemento A_{ij}. En esta notación el determinante de A se evalúa como la suma de productos de cada elemento de algún renglón o columna de A multiplicado por el cofactor de ese elemento. Entonces $\det(A)$ se expresa en términos de n determinantes de matrices de $(n-1) \times (n-1)$. Estos determinantes se evalúan luego en términos de determinantes de matrices de $(n-2) \times (n-2)$ y así sucesivamente, hasta que se obtienen matrices de 2×2. Los determinantes de las matrices de 2×2 se evalúan entonces como en el inciso 2.

Consideremos algunos ejemplos de esta técnica al evaluar el determinante de la matriz de 4×4

$$A = \begin{pmatrix} 2 & 1 & 1 & 5 \\ 1 & 1 & -4 & -1 \\ 2 & 0 & -3 & 1 \\ 3 & 6 & 1 & 2 \end{pmatrix}.$$

Primero evaluaremos el determinante de A por expansión sobre el cuarto renglón. Esto requiere de que conozcamos el cofactor de cada elemento del renglón. El cofactor de $A_{41} = 3$ es

$$(-1)^{4+1} \det\begin{pmatrix} 1 & 1 & 5 \\ 1 & -4 & -1 \\ 0 & -3 & 1 \end{pmatrix}.$$

Evaluemos el determinante anterior por expansión sobre la primera columna. Entonces

$$\det\begin{pmatrix}1 & 1 & 5\\ 1 & -4 & -1\\ 0 & -3 & 1\end{pmatrix} = (-1)^{1+1}(1)\det\begin{pmatrix}-4 & -1\\ -3 & 1\end{pmatrix} + (-1)^{2+1}(1)\det\begin{pmatrix}1 & 5\\ -3 & 1\end{pmatrix}$$

$$+ (-1)^{3+1}(0)\det\begin{pmatrix}1 & 5\\ -4 & -1\end{pmatrix}$$

$$= 1(1)[(-4)(1) - (-1)(-3)]$$

$$+ (-1)(1)[(1)(1) - (5)(-3)] + 0$$

$$= -7 - 16 + 0 = -23.$$

Así, el cofactor de A_{41} es $(-1)^5(-23) = 23$. Análogamente, el cofactor de $A_{42} = 6$ es

$$(-1)^{4+2}\det\begin{pmatrix}2 & 1 & 5\\ 1 & -4 & -1\\ 2 & -3 & 1\end{pmatrix}.$$

Evaluando este determinante sobre el segundo renglón da

$$\det\begin{pmatrix}2 & 1 & 5\\ 1 & -4 & -1\\ 2 & -3 & 1\end{pmatrix} = (-1)^{2+1}(1)\det\begin{pmatrix}1 & 5\\ -3 & 1\end{pmatrix} + (-1)^{2+2}(-4)\det\begin{pmatrix}2 & 5\\ 2 & 1\end{pmatrix}$$

$$+ (-1)^{2+3}(-1)\det\begin{pmatrix}2 & 1\\ 2 & -3\end{pmatrix}$$

$$= (-1)(1)[(1)(1) - (5)(-3)] + (1)(-4)[(2)(1) - (5)(2)]$$

$$+ (-1)(-1)[(2)(-3) - (1)(2)]$$

$$= -16 + 32 - 8 = 8.$$

Y así el cofactor de A_{42} es $(-1)^6(8) = 8$. El cofactor de $A_{43} = 1$ es

$$(-1)^{4+3}\det\begin{pmatrix}2 & 1 & 5\\ 1 & 1 & -1\\ 2 & 0 & 1\end{pmatrix}.$$

Si evaluamos este determinante por expansión sobre el tercer renglón, encontramos que

$$\det\begin{pmatrix}2 & 1 & 5\\ 1 & 1 & -1\\ 2 & 0 & 1\end{pmatrix} = (-1)^{3+1}(2)\det\begin{pmatrix}1 & 5\\ 1 & -1\end{pmatrix} + (-1)^{3+2}(0)\det\begin{pmatrix}2 & 5\\ 1 & -1\end{pmatrix}$$

$$+ (-1)^{3+3}(1)\det\begin{pmatrix}2 & 1\\ 1 & 1\end{pmatrix}$$

$$= 1(2)[(1)(-1) - (5)(1)] + 0 + 1(1)[(2)(1) - (1)(1)]$$

$$= -12 + 0 + 1 = -11.$$

Por tanto, el cofactor de A_{43} es $(-1)^7(-11) = 11$. Finalmente, el cofactor de $A_{44} = 2$ es

$$(-1)^{4+4}\det\begin{pmatrix}2 & 1 & 1\\ 1 & 1 & -4\\ 2 & 0 & -3\end{pmatrix}.$$

Calculando este determinante por expansión sobre la segunda columna, obtenemos

$$\det\begin{pmatrix}2 & 1 & 1\\ 1 & 1 & -4\\ 2 & 0 & -3\end{pmatrix} = (-1)^{1+1}(1)\det\begin{pmatrix}1 & -4\\ 2 & -3\end{pmatrix} + (-1)^{2+2}(1)\det\begin{pmatrix}2 & 1\\ 2 & -3\end{pmatrix}$$

$$+ (-1)^{3+2}(0)\det\begin{pmatrix}2 & 1\\ 1 & -4\end{pmatrix}$$

$$= (-1)(1)[(1)(-3) - (-4)(2)] + 1(1)[(2)(-3) - (1)(2)]$$

$$+ 0$$

$$= -5 - 8 + 0 = -13.$$

Por tanto el cofactor de A_{44} es $(-1)^8(-13) = -13$. Ahora podemos evaluar el determinante de A multiplicando cada elemento del cuarto renglón por su cofactor; esto da

$$\det(A) = 3(23) + 6(8) + 1(11) + 2(-13) = 102.$$

A fin de comparar calcularemos también el determinante de A por expansión sobre la segunda columna. El lector deberá verificar que los cofactores de A_{12}, A_{22} y A_{42} son 14, 40 y 8, respectivamente. Entonces

$$\det(A) = (-1)^{1+2}(1)\det\begin{pmatrix}1 & -4 & -1\\ 2 & -3 & 1\\ 3 & 1 & 2\end{pmatrix} + (-1)^{2+2}(1)\det\begin{pmatrix}2 & 1 & 5\\ 2 & -3 & 1\\ 3 & 1 & 2\end{pmatrix}$$

$$+ (-1)^{3+2}(0)\det\begin{pmatrix}2 & 1 & 5\\ 1 & -4 & -1\\ 3 & 1 & 2\end{pmatrix} + (-1)^{4+2}(6)\det\begin{pmatrix}2 & 1 & 5\\ 1 & -4 & -1\\ 2 & -3 & 1\end{pmatrix}$$

$$= 14 + 40 + 0 + 48 = 102.$$

Por supuesto, el hecho de que el valor 102 se haya obtenido de nuevo no es ninguna sorpresa puesto que el valor del determinante de A es independiente de la elección del renglón o columna utilizada en la expansión.

Obsérvese que el cálculo de $\det(A)$ es más fácil cuando se expande sobre la segunda columna que cuando se expande sobre el cuarto renglón. La diferencia es la presencia de un cero en la segunda columna, lo que hizo innecesario evaluar uno de los cofactores (el cofactor de A_{32}). Por esta razón es benéfico evaluar el determinante de la matriz expandiendo sobre el renglón o columna que tenga el mayor número de elementos nulos.

De hecho, es a menudo útil introducir ceros en la matriz por medio de operaciones elementales en los renglones antes de calcular el determinante. Esta técnica utiliza las tres primeras propiedades de los determinantes.

Propiedades del determinante

1. Si B es una matriz obtenida al intercambiar dos renglones o columnas de A, entonces $\det(B) = -\det(A)$.
2. Si B es una matriz obtenida al multiplicar todo elemento de algún renglón o columna de A por algún escalar c, entonces $\det(B) = c \times \det(A)$.
3. Si B es una matriz obtenida a partir de A al sumar un múltiplo del renglón i al renglón j o un múltiplo de la columna i a la columna j, donde $i \neq j$, entonces $\det(B) = \det(A)$.

Ilustraremos el uso de estas tres propiedades en la evaluación de determinantes calculando el determinante de la matriz A de 4×4 considerada anteriormente. Nuestro procedimiento será el de introducir ceros en la segunda columna de A utilizando la propiedad 3 y luego expandiendo sobre esa columna. (Las operaciones elementales sobre los renglones utilizadas consisten en sumar múltiplos del renglón 1 a los renglones 2 y 4.) Este proceso da

$$\det(A) = \det\begin{pmatrix} 2 & 1 & 1 & 5 \\ 1 & 1 & -4 & -1 \\ 2 & 0 & -3 & 1 \\ 3 & 6 & 1 & 2 \end{pmatrix} = \det\begin{pmatrix} 2 & 1 & 1 & 5 \\ -1 & 0 & -5 & -6 \\ 2 & 0 & -3 & 1 \\ -9 & 0 & -5 & -28 \end{pmatrix}$$

$$= 1(-1)^{1+2} \det\begin{pmatrix} -1 & -5 & -6 \\ 2 & -3 & 1 \\ -9 & -5 & -28 \end{pmatrix}.$$

El determinante resultante de una matriz de 3×3 puede ser evaluado de la misma manera. Utilizaremos operaciones elementales del tipo 3 en los renglones para introducir dos ceros en la primera columna y luego expandir sobre ella. Continuando con lo anterior tenemos

$$\det(A) = (-1)\cdot\det\begin{pmatrix} -1 & -5 & -6 \\ 2 & -3 & 1 \\ -9 & -5 & -28 \end{pmatrix} = (-1)\cdot\det\begin{pmatrix} -1 & -5 & -6 \\ 0 & -13 & -11 \\ 0 & 40 & 26 \end{pmatrix}$$

$$= (-1)\left[(-1)^{1+1}(-1)\det\begin{pmatrix} -13 & -11 \\ 40 & 26 \end{pmatrix}\right]$$

$$= (-13)(26) - (-11)(40) = 102.$$

El lector debiera comparar este cálculo de $\det(A)$ con los anteriores para determinar cuánto trabajo de menos se requirió cuando se emplearon las propiedades 1, 2 y 3.

En los siguientes capítulos tendremos a menudo que evaluar el determinante de matrices de formas especiales. Las siguientes tres propiedades de determinantes serán de gran utilidad para ello.

4. $\det(I) = 1$.
5. Si dos renglones (o columnas) de una matriz son idénticos, el determinante de la matriz es cero.
6. El determinante de una matriz triangular es igual al producto de los elementos de la diagonal.

Como ilustración de la propiedad 6, véase que

$$\det\begin{pmatrix} -3 & 0 & 0 \\ 1 & 4 & 0 \\ 2 & 5 & -6 \end{pmatrix} = (-3)(4)(-6) = 72.$$

Las cuatro propiedades restantes del determinante se utilizarán frecuentemente en capítulos posteriores. De hecho probablemente la propiedad más significativa del determinante es que proporciona una caracterización sencilla de las matrices invertibles (véase propiedad 10).

7. Para cualquier A, $\det(A) = \det(A^t)$.
8. Para cualquier $A, B \in M_{n\times n}(F)$, $\det(AB) = \det(A)\cdot\det(B)$.
9. Si Q es una matriz invertible, entonces $\det(Q^{-1}) = [\det(Q)]^{-1}$.
10. Una matriz Q es invertible si y sólo si $\det(Q) \neq 0$.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

   (a) El determinante de una matriz cuadrada puede ser calculado expandiendo la matriz sobre cualquier renglón o columna.
   (b) Al evaluar el determinante de una matriz, es conveniente expandir sobre un renglón o columna que contenga el mayor número de ceros.
   (c) Si dos renglones o columnas de A son idénticos, entonces $\det(A) = 0$.
   (d) Si B es una matriz obtenida al intercambiar dos renglones o columnas de A, entonces $\det(B) = \det(A)$.
   (e) Si B es la matriz obtenida al multiplicar todos los elementos de un renglón o columna de A por un escalar; $\det(B) = \det(A)$.
   (f) Si B es una matriz obtenida a partir de A sumando un múltiplo de algún renglón a un renglón distinto (o un múltiplo de alguna columna a alguna columna distinta), entonces $\det(B) = \det(A)$.
   (g) El determinante de una matriz triangular de $n \times n$ es igual al producto de sus elementos de la diagonal.

(h) $\det(A^t) = -\det(A)$.
(i) Si $A, B \in M_{n\times n}(F)$, entonces $\det(AB) = \det(A)\cdot\det(B)$.
(j) Si Q es una matriz invertible, entonces $\det(Q^{-1}) = [\det(Q)]^{-1}$.
(k) Una matriz Q es invertible si y sólo si $\det(Q) \neq 0$.

2. Evaluar el determinante de las siguientes matrices de 2×2.

(a) $\begin{pmatrix} 4 & -5 \\ 2 & 3 \end{pmatrix}$ (b) $\begin{pmatrix} -1 & 7 \\ 3 & 8 \end{pmatrix}$

(c) $\begin{pmatrix} 2+i & -1+3i \\ 1-2i & 3-i \end{pmatrix}$ (d) $\begin{pmatrix} 3 & 4i \\ -6i & 2i \end{pmatrix}$

3. Evaluar el determinante de las siguientes matrices en la manera indicada.

(a) Expandir sobre la segunda columna

$$\begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & -3 \\ 2 & 3 & 0 \end{pmatrix}$$

(b) Expandir sobre el tercer renglón

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 5 \\ -1 & 3 & 0 \end{pmatrix}$$

(c) Expandir sobre la primera columna

$$\begin{pmatrix} 0 & 1+i & 2 \\ -2i & 0 & 1-i \\ 3 & 4i & 0 \end{pmatrix}$$

(d) Expandir sobre el primer renglón

$$\begin{pmatrix} i & 2+i & 0 \\ -1 & 3 & 2i \\ 0 & -1 & 1-i \end{pmatrix}$$

(e) Expandir sobre la cuarta columna

$$\begin{pmatrix} 0 & 2 & 1 & 3 \\ 1 & 0 & -2 & 2 \\ 3 & -1 & 0 & 1 \\ -1 & 1 & 2 & 0 \end{pmatrix}$$

4. Evaluar el determinante de las siguientes matrices por cualquier método permitido.

(a) $\begin{pmatrix} 2 & 5 & 0 \\ -6 & 1 & 3 \\ 0 & -4 & 2 \end{pmatrix}$ (b) $\begin{pmatrix} -1 & 3 & 2 \\ 4 & -1 & 1 \\ 2 & 2 & 5 \end{pmatrix}$ (c) $\begin{pmatrix} 1 & 0 & 3 \\ 2 & -1 & 4 \\ -3 & 2 & 1 \end{pmatrix}$

(d) $\begin{pmatrix} i & 2 & -1 \\ 3 & 1+i & 2 \\ -2i & 1 & 4-i \end{pmatrix}$ (e) $\begin{pmatrix} -1 & 2+i & 3 \\ 1-i & i & 1 \\ 3i & 2 & -1+i \end{pmatrix}$

(f) $\begin{pmatrix} 1 & 0 & -2 & 3 \\ -3 & 1 & 1 & 2 \\ 0 & 4 & -1 & 1 \\ 2 & 3 & 0 & 1 \end{pmatrix}$ (g) $\begin{pmatrix} 4 & 2+i & 2i & 5+2i \\ 0 & 1-i & 1 & 3-4i \\ 0 & 0 & 3i & 6 \\ 0 & 0 & 0 & 2 \end{pmatrix}$

5.* Trabajar sobre el Ejercicio 9 de la Sección 4.3.

INDICE DE LAS DEFINICIONES DEL CAPITULO 4

Capítulo 5

Diagonalización

Este capítulo trata el llamado "problema de la diagonalización". Dada una transformación lineal $T: V \rightarrow V$, donde V es un espacio vectorial dimensionalmente finito, buscaremos respuestas a las siguientes interrogantes:

1. ¿Existe una base ordenada β para V tal que $[T]_\beta$ sea una matriz diagonal?
2. Si dicha base existe, ¿cómo puede encontrarse?

Como en general los cálculos que involucran a las matrices diagonales son sencillos, una respuesta afirmativa a la pregunta 1 nos conducirá a un mayor entendimiento de cómo la transformación T opera sobre V, y una respuesta a la pregunta 2 nos permitirá obtener soluciones fáciles a muchos problemas de orden práctico que pueden formularse dentro del contexto del álgebra lineal. Consideraremos algunos de estos problemas y sus soluciones dentro de este mismo capítulo —véase por ejemplo la Sección 5.3.

Una solución al problema de la diagonalización conduce de una manera natural a los conceptos de "eigenvalor" (valor propio o característico) y "eigenvector" (vector propio o característico). Aparte del importante papel que estos conceptos juegan en el problema de la diagonalización, su utilidad quedará también demostrada como valiosas herramientas en el estudio de muchas transformaciones no diagonalizables, tal como lo veremos en el Capítulo 6.

5.1 EIGENVALORES Y EIGENVECTORES

Como el problema de la diagonalización implica el estudio de una transformación que mapee a un espacio vectorial en sí mismo, es útil dar un nombre a tal transformación. En consecuencia, llamaremos a la transformación lineal $T: V \rightarrow V$ sobre un espacio vectorial V, un *operador lineal* sobre V.

Para un operador lineal dado T sobre un espacio vectorial dimensionalmente finito V, nos interesarán las matrices que representen a T de acuerdo con las diferentes bases ordenadas para V.

A lo largo de este capítulo omitiremos en general la palabra "ordenada" en la expresión "base ordenada".

Considérese un operador lineal T en un espacio vectorial dimensionalmente finito V y cualquier par de bases β y β' para V. Recuérdese del corolario al Teorema 2.27 que las matrices $[T]_\beta$ y $[T]_{\beta'}$ están relacionadas mediante la expresión

$$[T]_{\beta'} = Q^{-1}[T]_\beta Q,$$

donde Q es la matriz de cambio de coordenadas que transforma las coordenadas de β' en coordenadas de β. En la Sección 2.5 definimos tales matrices como matrices *similares*. Un caso especial de utilidad de este tipo de relaciones se demuestra en el siguiente teorema.

Teorema 5.1. *Sea* $A \in M_{n\times n}(F)$ *y sea* $\gamma = \{x_1, x_2, \ldots, x_n\}$ *una base cualquiera para* F^n. *Entonces* $[L_A]_\gamma = Q^{-1}AQ$, *donde* Q *es una matriz de* $n \times n$ *en la que la columna* j *es* x_j $(j = 1, 2, \ldots, n)$.

DEMOSTRACIÓN. Sea β la base estándar para F^n. Se puede ver fácilmente que la matriz Q es la matriz de cambio de coordenadas que transforma las coordenadas de γ en coordenadas de β. Por lo tanto

$$[L_A]_\gamma = Q^{-1}[L_A]_\beta Q = Q^{-1}AQ. \quad \blacksquare$$

Ejemplo 1. Para ilustrar el Teorema 5.1, sean

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \in M_{2\times 2}(R) \quad \text{y} \quad \gamma = \left\{ \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}.$$

Es muy sencillo verificar que si

$$Q = \begin{pmatrix} 1 & 1 \\ 3 & 2 \end{pmatrix},$$

entonces

$$Q^{-1} = \begin{pmatrix} -2 & 1 \\ 3 & -1 \end{pmatrix}$$

y por tanto

$$[L_A]_\gamma = Q^{-1}AQ = \begin{pmatrix} -2 & 1 \\ 3 & -1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 3 & 2 \end{pmatrix} = \begin{pmatrix} -11 & -8 \\ 18 & 13 \end{pmatrix}.$$

Como se mencionó anteriormente, las matrices que representan al mismo operador lineal relativo a bases diferentes son similares. Estableceremos en seguida el recíproco de este resultado.

Teorema 5.2. *Sea* T *un operador lineal en un espacio vectorial* V n-*dimensional, y sea* β *una base para* V. *Si* B *es cualquier matriz de* $n \times n$ *similar a* $[T]_\beta$, *entonces existe una base* β' *para* V *tal que* $B = [T]_{\beta'}$.

DEMOSTRACIÓN. Si B es similar a $[T]_\beta$, entonces existe una matriz invertible Q tal que $B = Q^{-1}[T]_\beta Q$. Supóngase que $\beta = \{x_1, x_2, \ldots, x_n\}$ y defínase

$$x'_j = \sum_{i=1}^{n} Q_{ij}x_i \qquad \text{para } 1 \leq j \leq n.$$

Entonces $\beta' = \{x'_1, x'_2, \ldots, x'_n\}$ es una base para V tal que Q es la matriz de cambio de coordenadas que transforma las coordenadas de β' en coordenadas de β. (Ejercicio 11 de la Sección 2.5.) Por lo tanto

$$[T]_{\beta'} = Q^{-1}[T]_\beta Q = B$$

de acuerdo con el corolario del Teorema 2.27. ■

El concepto de similitud es de utilidad en el estudio del problema de la diagonalización, pues puede ser utilizado para reformular el problema dentro del contexto matricial. Introduciremos ahora el concepto de diagonalizabilidad.

Definiciones. *Se dice que un operador lineal* T *sobre un espacio vectorial dimensionalmente finito* V *es* diagonalizable *si existe una base* β *para* V *tal que* $[T]_\beta$ *sea una matriz diagonal.*

Una matriz cuadrada A *es* diagonalizable *si* A *es similar a una matriz diagonal.*

El teorema siguiente relaciona estos dos conceptos y conduce a una reformulación del problema de la diagonalización dentro del contexto matricial.

Teorema 5.3. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V. *Los siguientes incisos son equivalentes:*

(a) T *es diagonalizable.*
(b) *Existe una base* β *para* V *tal que la matriz* $[T]_\beta$ *es diagonalizable.*
(c) *La matriz* $[T]_\gamma$ *es diagonalizable para cualquier base* γ *para* V.

DEMOSTRACIÓN. Si T es diagonalizable, entonces existe una base β para V tal que $[T]_\beta$ es una matriz diagonal. Entonces $[T]_\beta$ es trivialmente diagonalizable, por lo que (a) implica a (b).

Sea β una base para V tal que $[T]_\beta$ es diagonalizable y sea γ una base cualquiera para V. Entonces $[T]_\beta$ y $[T]_\gamma$ son similares. Luego, si $[T]_\beta$ es similar a una matriz diagonal, también $[T]_\gamma$ lo será de acuerdo con la transitividad de la relación de similitud. Y entonces $[T]_\gamma$ es diagonalizable, demostrando que (b) implica a (c).

Finalmente, si $[T]_\gamma$ es diagonalizable, existe una matriz diagonal D similar a $[T]_\gamma$. Luego, de acuerdo con el Teorema 5.2, existe una base β' para V tal que $[T]_{\beta'} = D$. Por tanto, T es diagonalizable y así (c) implica a (a). ■

Como una consecuencia inmediata de este Teorema tenemos el siguiente resultado de gran utilidad.

Corolario. *Una matriz A es diagonalizable si y sólo si L_A es diagonalizable.*

Como consecuencia del Teorema 5.3 podemos reformular el problema de la diagonalización de la manera siguiente.

1. ¿Es diagonalizable una matriz cuadrada A dada?
2. Si A es diagonalizable, ¿cómo puede determinarse una matriz Q invertible tal que $Q^{-1}AQ$ sea una matriz diagonal?

Presentaremos ahora el primero de los diferentes resultados que conducen a una solución del problema de la diagonalización.

Teorema 5.4. *Sea T un operador lineal en un espacio vectorial dimensionalmente finito V. Entonces T es diagonalizable si y sólo si existe una base $\beta = \{x_1, \ldots, x_n\}$ para V y escalares $\lambda_1, \ldots, \lambda_n$ (no necesariamente distintos) tales que $T(x_j) = \lambda_j x_j$, para $1 \leq j \leq n$. Bajo estas circunstancias*

$$[T]_\beta = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

DEMOSTRACIÓN. Supóngase que T es diagonalizable. Entonces existe una base β para V tal que $[T]_\beta = D$ es una matriz diagonal. Sean $\lambda_j = D_{jj}$ y $\beta = \{x_1, \ldots, x_n\}$. Entonces para cada j,

$$T(x_j) = \sum_{i=1}^{n} D_{ij}x_i = D_{jj}x_j = \lambda_j x_j.$$

Recíprocamente, supóngase que existe una base $\beta = \{x_1, \ldots, x_n\}$ y escalares $\lambda_1, \ldots, \lambda_n$ tales que $T(x_j) = \lambda_j x_j$. Entonces evidentemente

$$[T]_\beta = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$ ■

Del Teorema 5.4 se derivan las siguientes definiciones.

Definiciones. *Sea* T *un operador lineal en un espacio vectorial* V. *Un elemento no nulo* $x \in V$ *se llama* eigenvector *de* T *si existe un escalar* λ *tal que* $T(x) = \lambda x$. *Al escalar* λ *se le llama* eigenvalor *correspondiente al eigenvector* x.

Análogamente, si A *es una matriz de* $n \times n$ *en un campo* F, *un elemento no nulo* $x \in F^n$ *se denomina* eigenvector *de la matriz* A, *si* x *es un eigenvector de* L_A. *Como en el párrafo anterior, el escalar* λ *se denomina* eigenvalor *de* A *correspondiente al eigenvector* x.

A menudo se usan las palabras *vector característico* y *vector propio* en lugar de eigenvector. Los términos correspondientes para un eigenvalor son *valor característico* y *valor propio.*

Con esta terminología vemos que en el Teorema 5.4 la base β consta de eigenvectores de T y que los elementos de la diagonal de $[T]_\beta$ son los eigenvalores de T, por lo que el Teorema 5.4 puede ser enunciado de nuevo de la manera siguiente: *Un operador lineal* T *en un espacio vectorial dimensionalmente finito* V *es diagonalizable si y sólo si existe una base* β *para* V *compuesta por eigenvectores de* T. *Además, si* T *es diagonalizable,* $\beta = \{x_1, x_2, \ldots, x_n\}$ *es una base de eigenvectores de* T, *y* $D = [T]_\beta$; *entonces* D *es una matriz diagonal y* D_{ii} *es el eigenvalor correspondiente a* x_i $(i = 1, 2, \ldots, n)$.

Antes de continuar con nuestro análisis del problema de la diagonalización consideremos dos ejemplos que involucran eigenvectores y eigenvalores.

Ejemplo 2. Sea $C^\infty(R)$ el conjunto de todas las funciones $f\colon R \to R$ que tienen derivadas de todos los órdenes. (Por lo tanto $C^\infty(R)$ incluye a todas las funciones polinomiales, las funciones seno y coseno, las funciones exponenciales, etc.) Es fácil ver que $C^\infty(R)$ es un subespacio del espacio vectorial $\mathcal{F}(R, R)$ de todas las funciones de R en R como se definieron en la Sección 1.2. Defínase $T\colon C^\infty(R) \to C^\infty(R)$ mediante $T(y) = y'$, donde y' es la derivada de y. Puede verificarse fácilmente que T es un operador lineal en $C^\infty(R)$. Procederemos a determinar los eigenvalores y los eigenvectores de T.

Si λ es un eigenvalor de T, entonces existe un eigenvector $y \in C^\infty(R)$ tal que $y' = T(y) = \lambda y$. Esta es una ecuación diferencial de primer orden cuyas soluciones son de la forma $y(t) = ce^{\lambda t}$ para alguna constante c. En consecuencia, todo número real λ es un eigenvalor de T y los eigenvectores correspondientes son de la forma $ce^{\lambda t}$ para $c \neq 0$. (Nótese que si $\lambda = 0$, los eigenvectores son las funciones constantes no nulas.)

Ejemplo 3. Sea

$$A = \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix}, \quad x_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \text{y} \quad x_2 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}.$$

Como

$$\mathsf{L}_A(x_1) = \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \end{pmatrix} = -2\begin{pmatrix} 1 \\ -1 \end{pmatrix} = -2x_1,$$

x_1 es un eigenvector de L_A (y por tanto de A). También $\lambda_1 = -2$ es el eigenvalor asociado con x_1. Además,

$$\mathsf{L}_A(x_2) = \begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix}\begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 15 \\ 20 \end{pmatrix} = 5\begin{pmatrix} 3 \\ 4 \end{pmatrix} = 5x_2.$$

Luego, x_2 es un eigenvector de L_A (y de A) con $\lambda_2 = 5$ como eigenvalor asociado. Nótese que $\beta = \{x_1, x_2\}$ es una base para R^2, y por tanto, por el Teorema 5.4

$$[\mathsf{L}_A]_\beta = \begin{pmatrix} -2 & 0 \\ 0 & 5 \end{pmatrix}.$$

Finalmente, si

$$Q = \begin{pmatrix} 1 & 3 \\ -1 & 4 \end{pmatrix},$$

entonces por el Teorema 5.1,

$$Q^{-1}AQ = [\mathsf{L}_A]_\beta = \begin{pmatrix} -2 & 0 \\ 0 & 5 \end{pmatrix}.$$

El ejemplo anterior muestra una técnica para diagonalizar una matriz A de $n \times n$: Si $\beta = \{x_1, x_2, \ldots, x_n\}$ es una base para F^n que consta de los eigenvectores de A, y Q es la matriz de $n \times n$ cuya columna j es el eigenvector $x_j (j = 1, 2, \ldots, n)$, entonces $Q^{-1}AQ$ es una matriz diagonal. Para poder emplear este procedimiento necesitamos de un método para determinar los eigenvectores de una matriz u operador. Como se verá luego, los eigenvectores se determinan fácilmente una vez que se conocen los eigenvalores, por lo que principiaremos exponiendo un método para calcular los eigenvalores. Como ayuda en este cálculo utilizaremos el teorema siguiente para introducir el concepto de "determinante" de un operador lineal.

Teorema 5.5. *Sea T un operador lineal en un espacio vectorial dimensionalmente finito V y sean β y β' un par de bases cualesquiera para V. Entonces $det([\mathsf{T}]_\beta) = det([\mathsf{T}]_{\beta'})$.*

DEMOSTRACIÓN. Sean $A = [\mathsf{T}]_\beta$ y $B = [\mathsf{T}]_{\beta'}$. Como A y B son similares, existe una matriz invertible Q tal que $B = Q^{-1}AQ$. Por lo tanto

$$\begin{aligned} \det(B) &= \det(Q^{-1}AQ) = \det(Q^{-1}) \cdot \det(A) \cdot \det(Q) \\ &= [\det(Q)]^{-1} \cdot [\det(A)] \cdot [\det(Q)] = \det(A). \quad \blacksquare \end{aligned}$$

Este resultado da lugar a la siguiente definición.

Definición. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V. *Definimos el* determinante *de* T, *que denotaremos por* det(T), *de la manera siguiente: Escójase una base* β *para* V, *y defínase* det(T) = det ($[T]_\beta$). *Nótese que según el Teorema 5.5* det(T) *está bien definido, es decir, es independiente de la selección de la base* β.

Ejemplo 4. Sea T: $P_2(R) \to P_2(R)$, definida mediante $T(f) = f'$, la derivada de f. Para calcular det(T), sea $\beta = \{1, x, x^2\}$. Entonces β es una base para $P_2(R)$ y

$$[T]_\beta = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Por lo tanto $\det(T) = \det([T]_\beta) = 0$.

Nuestro siguiente resultado establece algunas propiedades del determinante de un operador lineal. Nótese la semejanza de estas propiedades con las que demostramos para el determinante de una matriz en el Capítulo 4.

Teorema 5.6. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente infinito* V. *Entonces*

(a) T *es invertible si y sólo si* $det(T) \neq 0$.
(b) *Si* T *es invertible, entonces* $det(T^{-1}) = [det(T)]^{-1}$.
(c) *Si* U: $V \to V$ *es lineal, entonces* $det(TU) = det(T) \cdot det(U)$.
(d) *Si* λ *es un escalar y* β *es una base cualquiera para* V, *entonces*

$$det(T - \lambda I_V) = det(A - \lambda I),$$

donde $A = [T]_\beta$.

DEMOSTRACIÓN. Las demostraciones de los incisos (a), (b) y (c) se dejan como ejercicios. Para demostrar el inciso (d), supóngase que λ es un escalar, β es una base para V y $A = [T]_\beta$. Entonces $[I_V]_\beta = I$, y por lo tanto $[T - \lambda I_V]_\beta = A - \lambda I$. Luego, por definición $\det(T - \lambda I_V) = \det(A - \lambda I)$. ■

El teorema siguiente nos proporciona un método para calcular los eigenvalores.

Teorema 5.7. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V *sobre un campo* F. *Un escalar* $\lambda \in F$ *es un eigenvalor de* T *si y sólo si* $det(T - \lambda I) = 0$.

DEMOSTRACIÓN. Supóngase que λ es un eigenvalor de T. Entonces existe un eigenvector $x \in V (x \neq 0)$ tal que $T(x) = \lambda x$. Luego $0 = T(x) - \lambda x = (T - \lambda I)(x)$. Como $x \neq 0$, $T - \lambda I$ no es invertible. Así, según el Teorema 5.6, $\det(T - \lambda I) = 0$.

Recíprocamente, supóngamos que $\det(T - \lambda I) = 0$. Entonces, de nuevo por el Teorema 5.6, $T - \lambda I$ no es invertible. Luego existe un vector no nulo $x \in V$ tal que $x \in N(T - \lambda I)$. Entonces $(T - \lambda I)(x) = 0$, y lógicamente $T(x) = \lambda x$. Por lo tanto x es un eigenvector (con λ como eigenvalor asociado) de T. ■

Corolario 1. *Sea* A *una matriz de* $n \times n$ *sobre un campo* F. *Entonces un escalar* $\lambda \in F$ *es un eigenvalor de* A *si y sólo si* $\det(A - \lambda I) = 0$.

DEMOSTRACIÓN. Ejercicio.

Ejemplo 5. Sea

$$A = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix} \in M_{2\times 2}(R).$$

Como

$$\det(A - \lambda I) = \det\begin{pmatrix} 1-\lambda & 1 \\ 4 & 1-\lambda \end{pmatrix} = \lambda^2 - 2\lambda - 3 = (\lambda - 3)(\lambda + 1),$$

los únicos eigenvalores de A son 3 y -1.

Ejemplo 6. Sea $T\colon P_2(R) \to P_2(R)$ el operador lineal definido mediante $T(f(x)) = f(x) + xf'(x) + f'(x)$, y sea $\beta = \{1, x, x^2\}$. Entonces β es una base para $P_2(R)$ y

$$[T]_\beta = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix}.$$

Como

$$\begin{aligned}\det(T - \lambda I) = \det([T]_\beta - \lambda I) &= \det\begin{pmatrix} 1-\lambda & 1 & 0 \\ 0 & 2-\lambda & 2 \\ 0 & 0 & 3-\lambda \end{pmatrix} \\ &= (1-\lambda)(2-\lambda)(3-\lambda) \\ &= -(\lambda-1)(\lambda-2)(\lambda-3),\end{aligned}$$

λ es un eigenvalor de T si y sólo si $\lambda = 1, 2$ o 3.

El Ejemplo 6 hace uso de la siguiente consecuencia evidente del Teorema 5.6.

Corolario 2. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V, *y sea* β *una base para* V. *Entonces* λ *es un eigenvalor de* T *si y sólo si es un eigenvalor de* $[T]_\beta$.

En los Ejemplos 5 y 6 el lector habrá podido observar que si A es una matriz de $n \times n$, entonces $\det(A - \lambda I_n)$ es un polinomio en λ de gra-

do n con un coeficiente principal $(-1)^n$. Los eigenvalores de A son sencillamente ceros de este polinomio; de manera que la siguiente definición es apropiada.

Definición. *Si* $A \in M_{n\times n}(F)$, *el polinomio* $det(A - tI_n)$ *en la incógnita* t *se denomina* polinomio característico *de* A.*

Se puede demostrar fácilmente que matrices semejantes tienen el mismo polinomio característico (ver Ejercicio 12). Este hecho permite la definición siguiente.

Definición. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V *con base* β. *Definimos al* polinomio característico $f(t)$ *de* T *como el polinomio característico de* $A = [T]_\beta$; *esto es,*

$$f(t) = det(A - tI).$$

La observación que precede a la definición muestra que ésta es independiente de la selección de la base β. A menudo representaremos al polinomio característico de un operador T mediante $\det(T - tI)$.

El siguiente resultado confirma nuestras observaciones sobre los Ejemplos 5 y 6; puede demostrarse mediante un argumento directo de inducción.

Teorema 5.8. *El polinomio característico de* $A \in M_{n\times n}(F)$ *es un polinomio de grado* n *con coeficiente principal* $(-1)^n$.

Las siguientes consecuencias del Teorema 5.8 son inmediatas. (Véase también el Corolario 2 del Teorema E.2.)

Corolario 1. *Sea* A *cualquier matriz de* $n \times n$ *y sea* $f(t)$ *el polinomio característico de* A. *Entonces*

(a) *Un escalar* λ *es un eigenvalor de* A *si y sólo si* λ *es un cero del polinomio* $f(t)$ *(es decir, si y sólo si* $f(\lambda) = 0$).
(b) A *tiene como máximo* n *eigenvalores distintos.*

Corolario 2. *Sea* T *un operador lineal en un espacio vectorial* n-*dimensional* V *con polinomio característico* $f(t)$. *Entonces*

(a) *Un escalar* λ *es un eigenvalor de* T *si y sólo si* λ *es un cero del polinomio* $f(t)$ *(es decir, si y sólo si* $f(\lambda) = 0$).
(b) T *tiene como máximo* n *eigenvalores distintos.*

---

* El lector observador debe haber notado que los elementos de la matriz $A - tI_n$ no son elementos del campo F. Sin embargo, son elementos de otro campo $F(t)$. (El campo $F(t)$ es el campo de los cocientes del anillo de los polinomios $F[t]$. Normalmente esto se estudia en cursos de álgebra abstracta.) En consecuencia los resultados sobre determinantes demostrados en el capítulo 4 continúan siendo ciertos dentro de este contexto.

Los dos corolarios anteriores nos proporcionan un método para determinar todos los eigenvalores de una matriz o de un operador, y nuestro siguiente resultado nos proporciona un procedimiento para determinar los eigenvectores correspondientes a un eigenvalor dado.

Teorema 5.9. *Sea* T *un operador lineal en un espacio vectorial* V, *y sea* λ *un eigenvalor de* T. *Un vector* $x \in V$ *es un eigenvector de* T *que corresponde a* λ *si y sólo si* $x \neq 0$ *y* $x \in N(T - \lambda I)$

DEMOSTRACIÓN. Ejercicio.

Ejemplo 7. Para encontrar todos los eigenvectores de la matriz

$$A = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix}$$

del Ejemplo 5, recuérdese que A tiene dos eigenvalores, $\lambda_1 = 3$ y $\lambda_2 = -1$. Principiaremos encontrando todos los eigenvectores correspondientes a $\lambda_1 = 3$. Sea

$$B = A - \lambda_1 I = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix} - \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 4 & -2 \end{pmatrix}.$$

Entonces

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in R^2$$

es un eigenvector correspondiente a $\lambda_1 = 3$ si y sólo si $x \neq 0$ y si $x \in N(L_B)$, esto es, $x \neq 0$ y

$$\begin{pmatrix} -2 & 1 \\ 4 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -2x_1 + x_2 \\ 4x_1 - 2x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Evidentemente el conjunto de todas las soluciones de la ecuación anterior es

$$\left\{ t\begin{pmatrix} 1 \\ 2 \end{pmatrix} : \quad t \in R \right\}.$$

Por lo tanto x es un eigenvector que corresponde a $\lambda_1 = 3$ si y sólo si

$$x = t\begin{pmatrix} 1 \\ 2 \end{pmatrix} \qquad \text{para alguna } t \neq 0.$$

Ahora, supóngase que x es un eigenvector de A que corresponde a $\lambda_2 = -1$. Sea

$$B = A - \lambda_2 I = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix} - \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix};$$

entonces

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in N(L_B)$$

si y sólo si x es una solución al sistema

$$\begin{cases} 2x_1 + x_2 = 0 \\ 4x_1 + 2x_2 = 0. \end{cases}$$

Por lo tanto

$$\mathsf{N}(\mathsf{L}_B) = \left\{ t\begin{pmatrix} 1 \\ -2 \end{pmatrix} : \quad t \in R \right\}$$

Luego, x es un eigenvector que corresponde a $\lambda_2 = -1$ si y sólo si

$$x = t\begin{pmatrix} 1 \\ -2 \end{pmatrix} \qquad \text{para alguna } t \neq 0.$$

Obsérvese que

$$\left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \end{pmatrix} \right\}$$

es una base para R^2 que está formada de eigenvectores de A. Luego, por el Teorema 5.4, L_A (y por lo tanto A) es diagonalizable. De hecho, si

$$Q = \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix},$$

el Teorema 5.1 implica que

$$Q^{-1}AQ = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}.$$

En el Ejemplo 6 vimos que el operador lineal T en $\mathsf{P}_2(R)$ definido mediante $\mathsf{T}(f(x)) = f(x) + xf'(x) + f'(x)$ tiene como eigenvalores a 1, 2 y 3. Ahora calcularemos los eigenvectores de T.

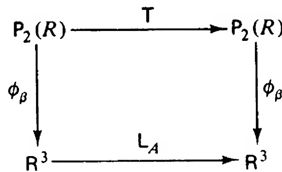


figura 5.1

Recordemos el diagrama de la Fig. 5.1 que procede de la Sección 2.4, donde $\beta = \{1, x, x^2\}$ y

$$A = [\mathsf{T}]_\beta = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix}.$$

Demostraremos que $v \in \mathsf{P}_2(R)$ es un eigenvector de T correspondiente a λ si y sólo si $\phi_\beta(v)$, es un eigenvector de A correspondiente a λ. (Este

argumento es válido para cualquier operador en un espacio vectorial dimensionalmente finito.) Si v es un eigenvector de T que corresponde a λ, entonces $\mathsf{T}(v) = \lambda v$. Por lo tanto

$$\mathsf{L}_A\phi_\beta(v) = \phi_\beta \mathsf{T}(v) = \phi_\beta(\lambda v) = \lambda\phi_\beta(v).$$

Ahora bien, $\phi_\beta(v) \neq 0$ puesto que ϕ_β es un isomorfismo. Luego $\phi_\beta(v)$ es un eigenvector de L_A (y por lo tanto de A) que corresponde a λ. Como el argumento anterior es reversible, podemos establecer de manera similar que si $\phi_\beta(v)$ es un eigenvector de A que corresponde a λ, entonces v es un eigenvector de T que corresponde a λ.

Una formulación equivalente del resultado demostrado en el párrafo anterior es que para cualquier eigenvalor λ de A (y por lo tanto de T), un vector $y \in \mathsf{R}^3$ es un eigenvector de A que corresponde a λ si y sólo si $\phi_\beta^{-1}(y)$ es un eigenvector de T correspondiente a λ. Este hecho nos permite calcular los eigenvectores de T tal como lo hicimos en el Ejemplo 7.

Sea $\lambda_1 = 1$ y defínase

$$B = A - \lambda_1 I = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{pmatrix}.$$

Puede demostrarse fácilmente que

$$\mathsf{N}(\mathsf{L}_B) = \left\{ a\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} : \quad a \in R \right\}.$$

Luego los eigenvectores de A que corresponden a λ_1 son de la forma

$$a\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

para alguna $a \neq 0$. En consecuencia, los eigenvectores de T que corresponden a $\lambda_1 = 1$ son de la forma

$$\phi_\beta^{-1}\left(a\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right) = a\phi_\beta^{-1}(e_1) = a$$

para alguna $a \neq 0$. Por lo tanto, los polinomios constantes no nulos son los eigenvectores de T que corresponden a λ_1.

Ahora sea $\lambda_2 = 2$ y defínase

$$B = A - \lambda_2 I = \begin{pmatrix} -1 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \end{pmatrix}.$$

De nuevo puede verificarse fácilmente que

$$\mathsf{N}(\mathsf{L}_B) = \left\{ a\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} : \quad a \in R \right\}.$$

Entonces, los eigenvectores de T que corresponden a λ_2 son de la forma

$$\phi_\beta^{-1}\left(a\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right) = a\phi_\beta^{-1}(e_1 + e_2) = a(1 + x) = a + ax$$

para alguna $a \neq 0$.

Finalmente, considérese $\lambda_3 = 3$ y

$$B = A - \lambda_3 I = \begin{pmatrix} -2 & 1 & 0 \\ 0 & -1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Como

$$\mathsf{N}(\mathsf{L}_B) = \left\{ a\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} : \quad a \in R \right\},$$

cualquier eigenvector de T correspondiente a $\lambda_3 = 3$ es de la forma

$$\phi_\beta^{-1}\left(a\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \right) = a\phi_\beta^{-1}(e_1 + 2e_2 + e_3) = a(1 + 2x + x^2) = a + 2ax + ax^2$$

para alguna $a \neq 0$.

Nótese también que $\gamma = \{1,\ 1 + x,\ 1 + 2x + x^2\}$ es una base para $\mathsf{P}_2(R)$ que consta de eigenvectores de T. Luego, T es diagonalizable y

$$[\mathsf{T}]_\gamma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Terminaremos esta sección analizando los eigenvectores y los eigenvalores desde un punto de vista geométrico. Si x es un eigenvector del operador lineal T en V, entonces $\mathsf{T}(x) = \lambda x$ para algún escalar λ. Sea $\mathsf{W} = L([x])$ el subespacio unidimensional de V generado por x. Si $y \in \mathsf{W}$, entonces $y = cx$ para algún escalar c. Entonces

$$\mathsf{T}(y) = \mathsf{T}(cx) = c\mathsf{T}(x) = c\lambda x = \lambda y \in \mathsf{W}.$$

De manera que T mapea a W en sí mismo. Si V es un espacio vectorial sobre el campo de los números reales, entonces W puede considerarse como

una recta que pasa por el origen de V (o sea, a través del cero). El operador T opera en los elementos de W multiplicando a cada elemento por el escalar λ. Existen diversas posibilidades para la acción de T dependiendo del valor de λ (véase Fig. 5.2).

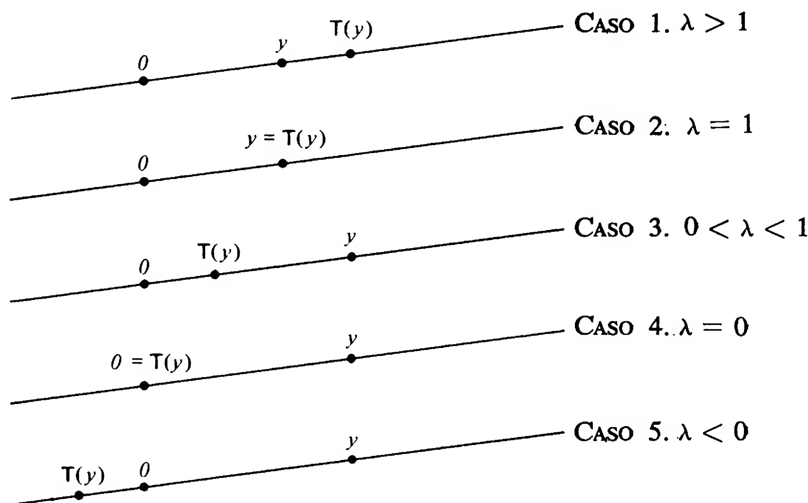
CASO 1. Si $\lambda > 1$, entonces T mueve a los elementos de W a puntos más lejanos al cero por un factor λ.

CASO 2. Si $\lambda = 1$, entonces T opera como la transformación identidad en W.

CASO 3. Si $0 < \lambda < 1$ T mueve a los elementos de W a puntos más cercanos a *0* por un factor λ.

CASO 4. Si $\lambda = 0$ entonces T opera como la transformación cero en W.

CASO 5. Si $\lambda < 0$ entonces T invierte la orientación de W; esto es, T desplaza los puntos de W de un lado del cero al otro.



La acción de T sobre $W = L(\{x\})$ cuando x es un eigenvector de T.

figura 5.2

Para ilustrar estas ideas, considérense los operadores lineales introducidos en los Ejemplos 6, 7 y 5 de la Sección 2.1. Recuérdese que el operador T: $R^2 \rightarrow R^2$ definido mediante $T(x_1, x_2) = (x_1, -x_2)$ es una reflexión sobre el eje x. Se ve fácilmente que T mapea a ambos ejes en sí mismos; luego e_1 y e_2 son eigenvectores de T (correspondientes respectiva-

mente a los eigenvalores 1 y -1). Obsérvese que T opera como la identidad sobre el eje x e invierte la orientación del eje y. Luego considérese la proyección sobre el eje de las x definida mediante $\mathsf{U}(x_1, x_2) = (x_1, 0)$. De nuevo es geométricamente evidente que U actúa como la identidad en el eje x y como la transformación cero en el eje y. Este comportamiento es una consecuencia del hecho de que e_1 y e_2 son eigenvectores de U correspondientes respectivamente a los eigenvalores 1 y 0. Finalmente, recuérdese que la rotación a través del ángulo θ es el operador $\mathsf{T}_\theta\colon \ \mathsf{R}^2 \to \mathsf{R}^2$ definido mediante $\mathsf{T}_\theta(x_1, x_2) = (x_1 \cos\theta - x_2 \operatorname{sen}\theta,\ x_1 \operatorname{sen}\theta + x_2 \cos\theta)$. Si $0 < \theta < \pi$ es geométricamente claro que T_θ no mapea a un espacio unidimensional de R^2 en sí mismo. Esta observación implica que T_θ no tiene eigenvectores (y por tanto tampoco eigenvalores). Para confirmar esta conclusión utilizando el Corolario 2 del Teorema 5.8, vemos que el polinomio característico de T_θ es

$$\det(\mathsf{T}_\theta - t\mathsf{I}) = \det\begin{pmatrix} \cos\theta - t & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta - t \end{pmatrix} = t^2 - (2\cos\theta)t + 1,$$

el cual no tiene ceros reales puesto que el discriminante $4\cos^2\theta - 4$ es negativo para $0 < \theta < \pi$. Luego, existen operadores (y por tanto matrices) sin eigenvalores ni eigenvectores. Por supuesto, tales operadores y matrices no son diagonalizables.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

   (a) Todo operador lineal en un espacio vectorial n-dimensional tiene n eigenvalores distintos.
   (b) Si una matriz real tiene un eigenvector, entonces tiene un número infinito de eigenvectores.
   (c) Existe una matriz cuadrada sin eigenvectores.
   (d) Los eigenvalores deben ser escalares no nulos.
   (e) Cualquier par de eigenvectores son linealmente independientes.
   (f) La suma de dos eigenvalores de un operador lineal T es también un eigenvalor de T.
   (g) Los operadores lineales de espacios vectoriales dimensionalmente infinitos nunca tienen eigenvalores.
   (h) Una matriz A de $n \times n$ con elementos de un campo F es similar a una matriz diagonal si y sólo si existe una base para F^n compuesta de eigenvectores de A.
   (i) Matrices similares siempre tienen los mismos eigenvalores.
   (j) Matrices similares siempre tienen los mismos eigenvectores.
   (k) La suma de dos eigenvectores de un operador T es siempre un eigenvector de T.

2. Para cada matriz A y base β, encontrar $[L_A]_\beta$. Encontrar también una matriz invertible Q tal que $[L_A]_\beta = Q^{-1}AQ$.

(a) $A = \begin{pmatrix} 1 & 3 \\ 1 & 1 \end{pmatrix}$ y $\beta = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$

(b) $A = \begin{pmatrix} 1 & 1 & -1 \\ 2 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ y $\beta = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} \right\}$

3. Para cada una de las siguientes matrices $A \in M_{n\times n}(F)$.

(i) Determínense todos los eigenvalores de A.
(ii) Para cada eigenvalor λ de A, encontrar el conjunto de eigenvectores correspondientes a λ.
(iii) De ser posible, encuéntrese una base para F^n compuesta por eigenvectores de A.
(iv) Si se tiene éxito en encontrar la base en (iii), determínese una matriz Q tal que $Q^{-1}AQ$ sea una matriz diagonal y calcúlese $Q^{-1}AQ$.

(a) $A = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}$ para $F = R$

(b) $A = \begin{pmatrix} 0 & -2 & -3 \\ -1 & 1 & -1 \\ 2 & 2 & 5 \end{pmatrix}$ para $F = R$

(c) $A = \begin{pmatrix} i & 1 \\ 2 & -i \end{pmatrix}$ para $F = C$

4. Sea $T: P_2(R) \to P_2(R)$ definida mediante $T((fx)) = f(x) + xf'(x)$. Encontrar todos los eigenvalores de T y encontrar una base β para $P_2(R)$ tal que $[T]_\beta$ sea una matriz diagonal.

5. Demostrar los incisos (a), (b) y (c) del Teorema 5.6.

6. Demostrar los Corolarios 1 y 2 del Teorema 5.7.

7. Demostrar el Teorema 5.9.

8. (a) Demostrar que un operador lineal T en un espacio vectorial dimensionalmente finito es invertible si y sólo si el cero no es un eigenvalor de T.
(b) Sea T un operador lineal invertible. Demostrar que un escalar λ es un eigenvalor de T si y sólo si λ^{-1} es un eigenvalor de T^{-1}.

9. Demostrar que los eigenvalores de una matriz triangular M son los elementos de la diagonal de M.

10. Sea V un espacio vectorial dimensionalmente finito y λ un escalar cualquiera.

(a) Para cualquier base β para V demostrar que $[\lambda I_V]_\beta = \lambda I$.
(b) Calcular el polinomio característico de λI_V.
(c) Demostrar que λI_V es diagonalizable y que tiene sólo un eigenvalor.

11. Una *matriz escalar* es una matriz cuadrada de la forma λI para algún escalar λ; o sea, una matriz escalar es una matriz diagonal en la cual todos los elementos de la diagonal son iguales.

(a) Demostrar que si A es similar a una matriz escalar λI, entonces $A = \lambda I$.
(b) Demostrar que una matriz diagonalizable que sólo tiene un eigenvalor es una matriz escalar.
(c) Concluir que la matriz

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

no es diagonalizable.

12. (a) Demostrar qué matrices similares tienen el mismo polinomio característico.
(b) Demostrar que la definición del polinomio característico de un operador lineal en un espacio vectorial dimensionalmente finito V es independiente de la selección de la base para V.

13. Demostrar las siguientes aseveraciones hechas en la página 241.

(a) Si $v \in P_2(R)$ y $\phi_\beta(v)$ es un eigenvector de A correspondiente al eigenvalor λ entonces v es un eigenvector de T que corresponde al eigenvalor λ.
(b) Si λ es un eigenvalor de A (y por tanto de T), entonces un vector $y \in R^3$ es un eigenvector de A correspondiente a λ si y sólo si $\phi_\beta^{-1}(y)$ es un eigenvector de T correspondiente a λ.

14.* Para cualquier matriz cuadrada A, demostrar que A y A^t tienen el mismo polinomio característico (y por tanto los mismos eigenvalores).

15.* (a) Sea T un operador lineal en un espacio vectorial V, y sea x un eigenvector de T correspondiente al eigenvalor λ. Para cualquier entero positivo m, demostrar que x es un eigenvector de T^m correspondiente al eigenvalor λ^m.
(b) Enunciar y demostrar el resultado para matrices, semejante al del inciso (a).

16. (a) Demostrar qué matrices similares tienen la misma traza. *Sugerencia:* Utilizar el Ejercicio 12 de la Sección 2.3.

(b) ¿Cómo se definiría la traza de un operador lineal en un espacio vectorial dimensionalmente finito? Justificar que la definición que se dé es correcta.

17. Sea $T: M_{n\times n}(F) \to M_{n\times n}(F)$ el mapeo definido mediante $T(A) = A^t$, la transpuesta de A.

(a) Verificar que T es un operador lineal en $M_{n\times n}(F)$.
(b) Demostrar que ± 1 son los únicos eigenvalores de T.
(c) Describir las matrices que sean eigenvectores correspondientes a los eigenvalores 1 y -1, respectivamente.

18. Demostrar que para cualesquiera A, $B \in M_{n\times n}(C)$ tal que B es invertible, existe un escalar $c \in C$ tal que $A + cB$ no es invertible. *Sugerencia:* Examinar a $\det(A + cB)$.

19.* Sean A y B matrices similares de $n \times n$. Demostrar que existe un espacio vectorial n-dimensional V, un operador lineal T en V y bases β y γ para V tales que $A = [T]_\beta$ y $B = [T]_\gamma$. *Sugerencia:* Utilizar el Ejercicio 12 de la Sección 2.5.

20. Sea A una matriz de $n \times n$ con polinomio característico

$$f(t) = (-1)^n t^n + a_{n-1}t^{n-1} + \ldots + a_1 t + a_0.$$

Demostrar que $f(0) = a_0 = \det(A)$. Deducir que A es invertible si y sólo si $a_0 \neq 0$.

21. Sean A y $f(t)$ como en el Ejercicio 20.

(a) Demostrar que $f(t) = (A_{11} - t)(A_{22} - t)\ldots(A_{nn} - t) + q(t)$, donde $q(t)$ es un polinomio en t de grado a lo más $(n - 2)$.
(b) Demostrar que $\mathrm{tr}(A) = (-1)^{n-1}a_{n-1}$.

22.* Sea T un operador lineal en un espacio vectorial dimensionalmente finito V sobre el campo F. Demostrar que si $g(t) \in P(F)$ y x es un eigenvector de T correspondiente al eigenvalor λ entonces $g(T)(x) = g(\lambda)x$.

5.2 DIAGONALIZABILIDAD

En la Sección 5.1 hemos presentado el problema de la diagonalización y vemos que no todos los operadores lineales ni todas las matrices son diagonalizables. Aun cuando fuimos capaces de diagonalizar ciertos operadores y matrices e incluso obtuvimos una condición necesaria y suficiente para diagonalizabilidad (Teorema 5.4), no hemos resuelto el problema de la diagonalización. Lo que aún se necesita es una prueba sencilla para determinar

si un operador o matriz puede ser diagonalizado y, si es posible, tener un algoritmo para obtener una base de eigenvectores. En esta sección desarrollaremos dicha prueba y un algoritmo.

En el Ejemplo 7 de la Sección 5.1 obtuvimos una base de eigenvectores escogiendo un eigenvector correspondiente a cada eigenvalor. En general, este procedimiento no proporciona una base, pero el teorema siguiente muestra que cualquier conjunto construido de esta manera debe ser linealmente independiente.

Teorema 5.10. *Sea* T *un operador lineal en* V *y sean* $\lambda_1, \lambda_2, \ldots, \lambda_k$ *eigenvalores de* T *diferentes. Si* $x_1, x_2, \ldots, x_k$ *son eigenvectores de* T *tales que* λ_j *corresponda a* $x_j (1 \leq j \leq k)$, *entonces* $\{x_1, x_2, \ldots, x_k\}$ *es linealmente independiente.*

DEMOSTRACIÓN. Utilizaremos inducción matemática sobre el número k. Supóngase que $k = 1$. Entonces $x_1 \neq 0$ ya que x_1 es un eigenvector, y entonces $\{x_1\}$ es linealmente independiente. Supóngase que el teorema se cumple siempre para $k - 1$ eigenvectores, donde $k - 1 \geq 1$ y que tenemos k eigenvectores $x_1, \ldots, x_k$ correspondientes a distintos eigenvalores $\lambda_1, \ldots, \lambda_k$. Deseamos demostrar que $\{x_1, \ldots, x_k\}$ es linealmente independiente. Supóngase que se tienen escalares $a_1, \ldots, a_k$ tales que

$$a_1x_1 + \ldots + a_kx_k = 0. \tag{1}$$

Aplicando T a ambos lados de la ecuación (1) obtenemos

$$a_1\mathsf{T}(x_1) + \ldots + a_k\mathsf{T}(x_k) = a_1\lambda_1x_1 + \ldots + a_k\lambda_kx_k = 0. \tag{2}$$

Ahora multiplicando ambos lados de la ecuación (1) por λ_k obtenemos

$$a_1\lambda_kx_1 + \ldots + a_k\lambda_kx_k = 0. \tag{3}$$

Luego, restando la ecuación (3) de la ecuación (2) tenemos

$$a_1(\lambda_1 - \lambda_k)x_1 + \ldots + a_{k-1}(\lambda_{k-1} - \lambda_k)x_{k-1} = 0.$$

Por la hipótesis de inducción $\{x_1, \ldots, x_{k-1}\}$ es linealmente independiente; por lo tanto

$$a_1(\lambda_1 - \lambda_k) = \ldots = a_{k-1}(\lambda_{k-1} - \lambda_k) = 0.$$

Como $\lambda_1, \ldots, \lambda_k$ son distintos, se tiene que $\lambda_i - \lambda_k \neq 0$ para $1 \leq i \leq k - 1$. Así $a_1 = \ldots = a_{k-1} = 0$, de manera que la ecuación (1) se reduce a $a_kx_k = 0$. Como $x_k \neq 0$, $a_k = 0$; por tanto, $a_1 = \ldots = a_k = 0$ y entonces $\{x_1, \ldots, x_k\}$ es linealmente independiente. ■

Corolario. *Sea* T *un operador lineal en* V, *un espacio vectorial dimensionalmente finito de dimensión* n. *Si* T *tiene* n *eigenvalores distintos, entonces* T *es diagonalizable.*

DEMOSTRACIÓN. Supóngase que T tiene n eigenvalores distintos $\lambda_1, \ldots, \lambda_n$ y sean $x_1, \ldots, x_n$ eigenvectores de T tales que λ_j corresponde a x_j

para $1 \leq j \leq n$. Por el Teorema 5.10 $\{x_1, \ldots, x_n\}$ es linealmente independiente y, como $\dim(V) = n$, este conjunto constituye una base para V. Luego, por el Teorema 5.4, T es diagonalizable. ■

Ejemplo 8. Sea

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \in M_{2\times 2}(R).$$

El polinomio característico de A (y por tanto de L_A) es

$$\det(A - tI) = \det\begin{pmatrix} 1-t & 1 \\ 1 & 1-t \end{pmatrix} = t(t-2),$$

y por lo tanto los eigenvalores de L_A son 0 y 2. Como L_A es un operador lineal en el espacio vectorial bidimensional R^2, concluimos del corolario anterior que L_A (y por tanto A) es diagonalizable.

Aun cuando el corolario del Teorema 5.10 proporciona una condición suficiente para la diagonalizabilidad, esta condición no es necesaria. De hecho, el operador identidad es diagonalizable, pero sólo tiene un eigenvalor, $\lambda = 1$.

Hemos visto que la existencia de eigenvalores es una condición necesaria para la diagonalizabilidad. El siguiente resultado nos dice más.

Teorema 5.11. *Sea* T *un operador lineal diagonalizable en un espacio vectorial* n*-dimensional* V, *y sea* $f(t)$ *el polinomio característico de* T. *Entonces* $f(t)$ *se descompone en un producto de* n *factores, todos de grado* 1; *esto es, existen escalares* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *(no necesariamente distintos) tales que*

$$f(t) = (-1)^n(t-\lambda_1)(t-\lambda_2)\ldots(t-\lambda_n).$$

DEMOSTRACIÓN. Supóngase que T es diagonalizable. Entonces existe una base β para V tal que $[T]_\beta = D$ es una matriz diagonal. Si

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

entonces

$$f(t) = \det(D - tI) = \det\begin{pmatrix} \lambda_1 - t & 0 & \cdots & 0 \\ 0 & \lambda_2 - t & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n - t \end{pmatrix}$$

$$= (\lambda_1 - t)(\lambda_2 - t)\cdots(\lambda_n - t) = (-1)^n(t-\lambda_1)(t-\lambda_2)\cdots(t-\lambda_n).$$ ■

De este teorema es claro que si T es un operador lineal diagonalizable en un espacio vectorial n-dimensional que no tiene n eigenvalores distintos, entonces el polinomio característico de T debe tener ceros múltiples. Esta observación nos conduce a la definición siguiente.

Definición. *Sea λ un eigenvalor de un operador lineal o de una matriz cuyo polinomio característico es* $f(t)$. *La* multiplicidad (algebraica) *de λ es el mayor entero positivo* k *para el que* $(t - \lambda)^k$ *es un factor de* $f(t)$.

Ejemplo 9. Sea

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 3 \\ 0 & 0 & 2 \end{pmatrix}.$$

Si $f(t)$ es el polinomio característico de A, entonces $f(t) = -(t-1)^2(t-2)$. Por tanto $\lambda = 1$ es un eigenvalor de A con multiplicidad 2 y $\lambda = 2$ es un eigenvalor de A con multiplicidad 1.

Si T es un operador lineal diagonalizable en un espacio vectorial V de dimensión finita, entonces existe una base β para V formada por eigenvectores de T. Sabemos del Teorema 5.4 que $[T]_\beta$ es una matriz diagonal en la que los elementos de la diagonal son los eigenvalores de T. Como el polinomio característico de T es $\det([T]_\beta - tI)$ se ve fácilmente que cada eigenvalor de T debe estar presente como un elemento de la diagonal de $[T]_\beta$ exactamente tantas veces como su multiplicidad. Por lo tanto β contiene tantos eigenvectores (linealmente independientes) correspondientes a un eigenvalor como la multiplicidad del mismo. Así vemos que el número de eigenvectores linealmente independientes correspondientes a un eigenvalor dado es muy importante para determinar cuándo un operador puede ser diagonalizado. Recordando del Teorema 5.9 que los eigenvectores de T correspondientes al eigenvalor λ son los vectores no nulos en el espacio nulo de $T - \lambda I$ es necesario el estudio de este conjunto de manera natural.

Definición. *Sea* T *un operador lineal en un espacio vectorial* V *y sea λ un eigenvalor de* T. *Defínase a* $E_\lambda = \{x \in V: T(x) = \lambda x\} = N(T - \lambda I_V)$. *El conjunto* E_λ *se denomina el* eigenespacio de T correspondiente al eigenvalor λ. *Como es de esperarse, por* eigenespacio *de una matriz* A *entendemos el eigenespacio correspondiente del operador* L_A.

Es claro que E_λ es un subespacio de V que contiene al vector cero y a los eigenvectores de T correspondientes al eigenvalor λ. El número de eigenvectores de T linealmente independientes correspondientes al eigenvalor λ es, por tanto, la dimensión de E_λ. Nuestro resultado siguiente relaciona esta dimensión con la multiplicidad de λ.

Teorema 5.12. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V. *Si* λ *es un eigenvalor de* T *de multiplicidad* m, *entonces* $1 \leq dim(E_\lambda) \leq m$.

DEMOSTRACIÓN. Tómese una base $\{x_1, \ldots, x_p\}$ para E_λ y extiéndase ésta a una base $\beta = \{x_1, \ldots, x_p, x_{p+1}, \ldots, x_n\}$ para V. Obsérvese que $x_i (1 \leq i \leq p)$ es un eigenvector de T que corresponde a λ y sea $A = [T]_\beta$. Entonces

$$A = \begin{pmatrix} B_1 & B_2 \\ O & B_3 \end{pmatrix},$$

donde $B_1 = \lambda I_p$ y O es la matriz cero.

Por el Ejercicio 9 de la Sección 4.3, el polinomio característico de T es

$$f(t) = \det(A - tI_n) = \det \begin{pmatrix} B_1 - tI_p & B_2 \\ O & B_3 - tI_{n-p} \end{pmatrix}$$
$$= \det(B_1 - tI_p) \cdot \det(B_3 - tI_{n-p}).$$

Sea $g(t) = \det(B_3 - tI_{n-p})$ el polinomio característico de B_3. Se ve claramente que $\det(B_1 - tI_p) = (\lambda - t)^p = (-1)^p(t - \lambda)^p$. Por lo tanto $f(t) = (-1)^p(t - \lambda)^p g(t)$, de manera que la multiplicidad de λ es al menos p. Pero $\dim(E_\lambda) = p$ por lo que $\dim(E_\lambda) \leq m$. ■

Ejemplo 10. Sea T: $P_2(R) \to P_2(R)$ el operador lineal definido mediante $T(f) = f'$, la derivada de f. La matriz de T con respecto a la base $\beta = \{1, x, x^2\}$ para $P_2(R)$ es

$$[T]_\beta = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Consecuentemente el polinomio característico de T es

$$\det([T]_\beta - tI) = \det \begin{pmatrix} -t & 1 & 0 \\ 0 & -t & 2 \\ 0 & 0 & -t \end{pmatrix} = -t^3.$$

Entonces T tiene solamente un eigenvalor ($\lambda = 0$) con multiplicidad 3. Luego $E_\lambda = N(T - \lambda I) = N(T)$. Por lo tanto E_λ es el subespacio de $P_2(R)$ que contiene a los polinomios constantes. Y en este caso $\{1\}$ es una base para E_λ y $\dim(E_\lambda) = 1$. Consecuentemente no existe una base para $P_2(R)$ que conste de eigenvectores de T, de modo que T no es diagonalizable.

Ejemplo 11. Sea T un operador lineal en R^3 definido mediante

$$T\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 4a_1 & & + a_3 \\ 2a_1 & + 3a_2 & + 2a_3 \\ a_1 & & + 4a_3 \end{pmatrix}.$$

Determinaremos el eigenespacio de T correspondiente a cada eigenvalor. Si β es la base ordinaria para R^3, entonces

$$[T]_\beta = \begin{pmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix}.$$

Por tanto el polinomio característico de T es

$$\det([T]_\beta - tI) = \det\begin{pmatrix} 4-t & 0 & 1 \\ 2 & 3-t & 2 \\ 1 & 0 & 4-t \end{pmatrix} = -(t-5)(t-3)^2.$$

De manera que los eigenvalores de T son $\lambda_1 = 5$ y $\lambda_2 = 3$ con multiplicidades 1 y 2, respectivamente.

Como

$$E_{\lambda_1} = N(T - \lambda_1 I) = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in R^3 \colon \begin{pmatrix} -1 & 0 & 1 \\ 2 & -2 & 2 \\ 1 & 0 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\},$$

E_λ es el espacio de soluciones del sistema de ecuaciones

$$\begin{cases} -x_1 \phantom{{}-2x_2} + x_3 = 0 \\ 2x_1 - 2x_2 + 2x_3 = 0 \\ x_1 \phantom{{}-2x_2} - x_3 = 0. \end{cases}$$

Se ve fácilmente (utilizando las técnicas del Capítulo 3) que

$$\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \right\}$$

es una base para E_{λ_1}. De donde $\dim(E_{\lambda_1}) = 1$.

De manera análoga $E_{\lambda_2} = N(T - \lambda_2 I)$ es el espacio de soluciones del sistema

$$\begin{cases} x_1 + x_3 = 0 \\ 2x_1 + 2x_3 = 0 \\ x_1 + x_3 = 0. \end{cases}$$

Y entonces

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$$

es una base para E_{λ_2} y $\dim(E_{\lambda_2}) = 2$.

En este caso la multiplicidad de cada eigenvalor λ_i es igual a la dimensión del eigenespacio correspondiente E_{λ_i}. Obsérvese que

$$\left\{\begin{pmatrix}1\\2\\1\end{pmatrix}, \begin{pmatrix}1\\0\\-1\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right\}$$

es una base para R^3 que consta de eigenvectores de T. En consecuencia T es diagonalizable.

Los Ejemplos 10 y 11 sugieren la siguiente conjetura: Si T es un operador lineal en un espacio vectorial dimensionalmente finito V tal que el polinomio característico de T se puede descomponer en un producto de factores de grado 1, entonces T es diagonalizable si y sólo si la multiplicidad de cada eigenvalor es igual a la dimensión del eigenespacio de T correspondiente a ese eigenvalor. Esta conjetura es, de hecho, cierta, pero su demostración implica una complicación que aún no estamos preparados para resolver. La dificultad es que aún no sabemos en general que la unión de las bases para cada uno de los eigenespacios será una base para V. (Este hecho estuvo claro dentro del contexto del Ejemplo 11 pero el caso general no ha sido demostrado.) Nótese que el Teorema 5.10 no es útil en este caso a menos que cada eigenespacio sea de dimensión 1. Así, debemos apartarnos un poco del problema de la diagonalización para establecer este hecho, el cual requerirá de la generalización del concepto de suma directa (tal como se definió en la Sección 1.3). Para este fin será conveniente expresar una suma de subespacios $W_1, W_2, \ldots, W_k$ (no necesariamente directa) como

$$\sum_{i=1}^{k} W_i.$$

Definición. *Sean* $W_1, W_2, \ldots, W_k$ *subespacios de un espacio vectorial* V. *Escribiremos* $V = W_1 \oplus W_2 \oplus \ldots \oplus W_k$ *y llamaremos a* V *la* suma directa *de* $W_1, W_2, \ldots, W_k$ *si*

$$V = \sum_{i=1}^{k} W_i$$

y

$$W_i \cap (\sum_{j \neq i} W_j) = \{0\} \qquad \textit{para cada } i (1 \leq i \leq k).$$

Ejemplo 12. Sea $V = R^4$ y sean $W_1 = \{(a, b, 0, 0)\colon a, b \in R\}$, $W_2 = \{(0, 0, c, 0)\colon c \in R\}$ y $W_3 = \{(0, 0, 0, d)\colon d \in R\}$. Para cualquier elemento (a, b, c, d) de V

$$(a, b, c, d) = (a, b, 0, 0) + (0, 0, c, 0) + (0, 0, 0, d) \in W_1 + W_2 + W_3.$$

Luego entonces

$$V = \sum_{i=1}^{3} W_i.$$

Para demostrar que V es la suma directa de W_1, W_2 y W_3 debemos de mostrar que $W_1 \cap (W_2 + W_3) = \{0\}$, $W_2 \cap (W_1 + W_3) = \{0\}$ y $W_3 \cap (W_1 + W_2) = \{0\}$. Pero estas igualdades son evidentes; de modo que $V = W_1 \oplus W_2 \oplus W_3$.

Nuestro resultado siguiente contiene varias condiciones que son equivalentes a la definición de suma directa. Nótese que este teorema contiene al Teorema 1.6 como un caso especial.

Teorema 5.13. *Sean* $W_1, W_2, \ldots, W_k$ *subespacios de un espacio vectorial dimensionalmente finito* V. *Las siguientes condiciones son equivalentes:*

(a) $V = W_1 \oplus W_2 \oplus \ldots \oplus W_k$.

(b) $V = \sum_{i=1}^{k} W_i$ *y, para vectores* $x_1, x_2, \ldots, x_k$ *cualesquiera tales que* $x_i \in W_i$ $(i = 1, 2, \ldots, k)$, *si* $x_1 + x_2 + \ldots + x_k = 0$, *entonces* $x_i = 0$ $(i = 1, 2, \ldots, k)$.

(c) *Cada vector* v *en* V *puede escribirse de manera única en la forma* $v = x_1 + x_2 + \ldots + x_k$, *donde* $x_i \in W_i (i = 1, 2, \ldots, k)$.

(d) *Si para toda* $i = 1, 2, \ldots, k$, γ_i *es una base ordenada cualquiera para* W_i, *entonces* $\gamma_1 \cup \gamma_2 \cup \ldots \cup \gamma_k$ * *es una base ordenada para* V.

(e) *Para toda* $i = 1, 2, \ldots, k$ *existe una base ordenada* γ_i *para* $W_i (i = 1, 2, \ldots, k)$ *tal que* $\gamma_1 \cup \gamma_2 \cup \ldots \cup \gamma_k$ *es una base ordenada para* V.

DEMOSTRACIÓN. Si (a) es cierta entonces, por definición,

$$V = \sum_{i=1}^{k} W_i.$$

Supóngase que $x_1, x_2, \ldots, x_k$ son vectores tales que $x_i \in W_i (i = 1, 2, \ldots, k)$ y $x_1 + x_2 + \ldots + x_k = 0$. Entonces para cualquier i

$$-x_i = \sum_{j \neq i} x_j \in \sum_{j \neq i} W_j.$$

Pero también

$$-x_i \in W_i, \text{ y así } -x_i \in W_i \cap (\sum_{j \neq i} W_j) = \{0\}.$$

Por lo tanto $x_i = 0$, lo que demuestra a (b).

Demostraremos en seguida que (b) implica a (c). Puesto que de acuerdo con (b)

$$V = \sum_{i=1}^{k} W_i$$

* Consideraremos a $\gamma_1 \cup \gamma_2 \cup \ldots \cup \gamma_k$ como una base ordenada de la manera normal —los vectores de γ_1 se enumeran primero (en el mismo orden que en γ_1), luego los vectores de γ_2 (en el mismo orden que en γ_2), etc.

cualquier vector $v \in V$ puede ser representado en la forma $v = x_1 + x_2 + \ldots + x_k$ para algunos elementos $x_i \in W_i (i = 1, 2, \ldots, k)$. Debemos demostrar que esta representación es única. Supóngase por tanto que $v = y_1 + y_2 + \ldots + y_k$, donde $y_i \in W_i (i = 1, 2, \ldots, k)$. Entonces

$$(x_1 - y_1) + (x_2 - y_2) + \ldots + (x_k - y_k) = 0.$$

Pero como $x_i - y_i \in W_i$, se deduce de (b) que $x_i - y_i = 0 (i = 1, 2, \ldots, k)$. Luego $x_i = y_i$ para cada i, lo que demuestra la unicidad de la representación.

Para demostrar que (c) implica a (d), sea γ_i una base para $W_i (i = 1, 2, \ldots, k)$. Como de acuerdo con (c)

$$V = \sum_{i=1}^{k} W_i.$$

es evidente que $\gamma_1 \cup \gamma_2 \cup \ldots \cup \gamma_k$ genera a V. Supóngase que existen vectores $x_{ij} \in \gamma_i (j = 1, 2, \ldots, m_i$ e $i = 1, 2, \ldots, k)$ y escalares a_{ij} tales que

$$\sum_{i,j} a_{ij} x_{ij} = 0.$$

Hágase

$$y_i = \sum_{j=1}^{m_i} a_{ij} x_{ij};$$

entonces $y_i \in L(\gamma_i) = W_i$ y

$$y_1 + y_2 + \ldots + y_k = \sum_{i,j} a_{ij} x_{ij} = 0.$$

Puesto que $0 \in W_i$ para toda i y $0 + 0 + \ldots + 0 = y_1 + y_2 + \ldots + y_k$, la condición (c) implica que $y_i = 0$ para toda i. Luego,

$$0 = y_i = \sum_{j=1}^{m_i} a^i{}_j x_{ij}$$

para toda i. Pero como γ_i es linealmente independiente, se obtiene que $a_{ij} = 0$ para $j = 1, 2, \ldots, m_i$ y toda i. Por lo tanto $\gamma_1 \cup \gamma_2 \cup \ldots \cup \gamma_k$ es linealmente independiente y entonces es una base para V.

Es inmediato que (d) implica a (e).

Finalmente, demostraremos que (e) implica a (a). Si γ_i es una base para $W_i (i = 1, 2, \ldots, k)$ tal que $\gamma_1 \cup \gamma_2 \cup \ldots \cup \gamma_k$ es una base para V, entonces

$$\begin{aligned} V &= L(\gamma_1 \cup \gamma_2 \cup \ldots \cup \gamma_k) \\ &= L(\gamma_1) + L(\gamma_2) + \ldots + L(\gamma_k) = \sum_{i=1}^{k} W_i \end{aligned}$$

mediante sucesivas aplicaciones del Ejercicio 12 de la Sección 1.4. Fíjese un índice i y supóngase que

$$0 \neq v \in W_i \cap (\sum_{j \neq i} W_j).$$

Entonces

$$v \in W_i = L(\gamma_i) \quad \text{y} \quad v \in \sum_{j \neq i} W_j = L(\bigcup_{j \neq i} \gamma_j).$$

Por lo tanto v es una combinación lineal no trivial de γ_i y $\bigcup_{j \neq i} \gamma_j$, de manera que v puede expresarse como combinación lineal de $\gamma_1 \cup \gamma_2 \cup \ldots \cup \gamma_k$ en más de una manera. Pero esas representaciones contradicen al Teorema 1.9, por lo que se concluye que

$$W_i \cap (\sum_{j \neq i} W_j) = \{0\},$$

demostrando (a). ■

La razón para discutir sumas directas es para permitirnos demostrar que si T es un operador lineal diagonalizable en un espacio vectorial dimensionalmente finito V, entonces la unión de las bases para cada uno de los eigenespacios de T es una base para V. El teorema anterior muestra que esta condición es equivalente a demostrar que, si T es diagonalizable, entonces V es la suma directa de sus eigenespacios. Ahora formalizaremos este importante resultado.

Teorema 5.14. *Sea* T *un operador lineal en un espacio vectorial* n*-dimensional* V. *Supóngase que el polinomio característico de* T *se puede descomponer en un producto de factores de grado 1 y sean* $\lambda_1, \lambda_2, \ldots, \lambda_k$ *los distintos eigenvalores de* T. *Entonces los siguientes incisos son equivalentes:*

(a) T *es diagonalizable.*

(b) $V = E_{\lambda_1} \oplus E_{\lambda_2} \oplus \ldots \oplus E_{\lambda_k}$.

(c) *Si* $d_j = dim(E_{\lambda_j})$ *para* $1 \leq j \leq k$, *entonces* $d_1 + d_2 + \ldots + d_k = n$.

(d) *Si* m_j *es la multiplicidad de* λ_j *para toda* $j (1 \leq j \leq k)$, *entonces* $dim(E_{\lambda_j}) = m_j (j = 1, 2, \ldots, k)$.

DEMOSTRACIÓN. Primero demostraremos que (a) implica a (b). Si T es diagonalizable, entonces V tiene una base que consiste en eigenvectores de T, de donde se deduce fácilmente que

$$V = \sum_{i=1}^{k} E_{\lambda_i}.$$

Sean $x_i \in E_{\lambda_i} (i = 1, 2, \ldots, k)$ vectores tales que $x_1 + x_2 + \ldots + x_k = 0$. Ahora bien, cada x_i es o bien el vector nulo o un eigenvector de T correspondiente a λ_i. Como por el Teorema 5.10 el conjunto de estos vectores

no nulos x_i es linealmente independiente, $x_1 + x_2 + \ldots + x_k = 0$ implica que $x_1 = x_2 = \ldots = x_k = 0$. Luego, por el Teorema 5.13 $V = E_{\lambda_1} \oplus \oplus E_{\lambda_2} \oplus \ldots \oplus E_{\lambda_k}$.

Si $V = E_{\lambda_1} \oplus E_{\lambda_2} \oplus \ldots \oplus E_{\lambda_k}$, entonces se infiere del Ejercicio 5 que

$$n = \dim(V) = \sum_{i=1}^{k} \dim(E_{\lambda_i}) = d_1 + d_2 + \ldots + d_k.$$

Así (b) implica a (c).

Demostraremos ahora que (c) implica a (d). Supóngase que

$$\sum_{i=1}^{k} d_i = n.$$

Por el Teorema 5.12, $d_j \leq m_j$ para toda j y por tanto

$$n = \sum_{i=1}^{k} d_i \leq \sum_{i=1}^{k} m_i.$$

Pero

$$\sum_{i=1}^{k} m_i = n$$

puesto que el polinomio característico se descompone en un producto de factores de grado 1. Luego, ya que

$$\sum_{i=1}^{k} (m_i - d_i) = 0 \quad \text{y} \quad m_i - d_i \geq 0$$

para cada i, podemos concluir que $d_i = m_i$ para cada i.

Finalmente, demostraremos que (d) implica a (a). Supóngase que $d_j = \dim(E_{\lambda_j}) = m_j$ para toda j y sea

$$W = \sum_{i=1}^{k} E_{\lambda_i}.$$

Un argumento similar al del primer párrafo de esta demostración muestra que $W = E_{\lambda_1} \oplus E_{\lambda_2} \oplus \ldots \oplus E_{\lambda_k}$. Si β_i es una base ordenada para $E_{\lambda_i} (i = = 1, 2, \ldots, k)$, entonces de acuerdo con el Teorema 5.13 $\beta_1 \cup \beta_2 \cup \cup \ldots \cup \beta_k$ es una base para W. Pero $\beta_1 \cup \beta_2 \cup \ldots \cup \beta_k$ contiene

$$\sum_{i=1}^{k} \dim(E_{\lambda_i}) = \sum_{i=1}^{k} m_i = n$$

vectores y por lo tanto $W = V$. De modo que V tiene una base $\beta_1 \cup \cup \beta_2 \cup \ldots \cup \beta_k$ que está formada por eigenvectores de T. Luego T es diagonalizable, lo que demuestra (a). ■

Este teorema completa nuestro estudio del problema de la diagonalización. Resumiremos algunos de nuestros resultados anteriores en la prueba y en el algoritmo siguientes.

Una prueba para diagonalizabilidad

Sea T un operador lineal en un espacio vectorial n-dimensional. Entonces T es diagonalizable si y sólo si se satisfacen las siguientes dos condiciones.

1. El polinomio característico de T se descompone en un producto de factores de grado 1.
2. La multiplicidad de λ es igual a $n - \text{rango}(\mathsf{T} - \lambda \mathsf{I})$ para cada eigenvalor λ de T.

Obsérvese que la condición 2 queda automáticamente satisfecha para eigenvalores de multiplicidad 1 (Teorema 5.12). Luego, la condición 2 sólo debe verificarse para eigenvalores de multiplicidad mayor que 1.

Un algoritmo para la diagonalización

Sea T un operador lineal diagonalizable en un espacio vectorial dimensionalmente finito V y sean $\lambda_1, \ldots, \lambda_k$ los distintos eigenvalores de T. Para cada j, sea β_j una base para $\mathsf{E}_{\lambda_j} = \mathsf{N}(\mathsf{T} - \lambda_j \mathsf{I})$ y sea $\beta = \beta_1 \cup \beta_2 \cup \cup \ldots \cup \beta_k$. Entonces β es una base para V, y $[\mathsf{T}]_\beta$ es una matriz diagonal.

Ejemplo 13. Probaremos si la matriz

$$A = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix} \in \mathsf{M}_{3\times 3}(R)$$

es diagonalizable. Como la prueba anterior está enunciada para operadores lineales en vez de para matrices, aplicaremos la prueba al operador L_A.

El polinomio característico de L_A es $\det(A - tI) = -(t-4)(t-3)^2$. Por lo tanto L_A tiene como eigenvalores a $\lambda_1 = 4$ y $\lambda_2 = 3$ con multiplicidades respectivas 1 y 2. Claramente se satisface la condición 1 de la prueba de diagonalizabilidad y como λ_1 tiene multiplicidad 1, la condición 2 se satisface para λ_1 por lo que sólo tenemos que verificar la condición 2 de la prueba para λ_2. Como

$$B = A - \lambda_2 I = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

tiene rango 2, $3 - \text{rango}(B) = 1$. Así, la condición 2 de la prueba falla para λ_2 y en consecuencia L_A y (por tanto A) no es diagonalizable.

Ejemplo 14. Sea $T: R^3 \to R^3$ definida mediante

$$T\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} -2b - 3c \\ a + 3b + 3c \\ c \end{pmatrix}.$$

Probaremos si T es diagonalizable. Siendo γ la base estándar para R^3, tenemos

$$[T]_\gamma = \begin{pmatrix} 0 & -2 & -3 \\ 1 & 3 & 3 \\ 0 & 0 & 1 \end{pmatrix}.$$

El polinomio característico de T es $-(t-1)^2(t-2)$. Luego T tiene 2 eigenvalores: $\lambda_1 = 1$ con multiplicidad 2 y $\lambda_2 = 2$ con multiplicidad 1. Nótese que la condición 1 de la prueba para diagonalizabilidad queda satisfecha. Ahora consideraremos la condición 2.

Para $\lambda_1 = 1$,

$$3 - \text{rango}(T - \lambda_1 I) = 3 - \text{rango}\begin{pmatrix} -1 & -2 & -3 \\ 1 & 2 & 3 \\ 0 & 0 & 0 \end{pmatrix} = 3 - 1 = 2.$$

Luego la dimensión de E_{λ_1} es la misma que la multiplicidad de λ_1. Como λ_2 tiene multiplicidad 1, la dimensión de E_{λ_2} es igual a la multiplicidad de λ_2. Por tanto T es diagonalizable.

Encontraremos ahora una base β para R^3 tal que $[T]_\beta$ sea una matriz diagonal. Dado que

$$E_{\lambda_1} = N(T - \lambda_1 I) = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in R^3 : \begin{pmatrix} -1 & -2 & -3 \\ 1 & 2 & 3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0 \right\},$$

E_{λ_1} es el conjunto de soluciones de

$$\begin{cases} -x_1 - 2x_2 - 3x_3 = 0 \\ x_1 + 2x_2 + 3x_3 = 0, \end{cases}$$

que tiene como base a

$$\beta_1 = \left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

También

$$E_{\lambda_2} = N(T - \lambda_2 I) = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in R^3 : \begin{pmatrix} -2 & -2 & -3 \\ 1 & 1 & 3 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0 \right\}.$$

Luego E_{λ_2} es el conjunto de soluciones de

$$\begin{cases} -2x_1 - 2x_2 - 3x_3 = 0 \\ \quad x_1 + \quad x_2 + 2x_3 = 0 \\ \qquad\qquad\quad - \quad x_3 = 0, \end{cases}$$

que tiene como base a

$$\beta_2 = \left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}$$

Sea $\beta = \beta_1 \cup \beta_2$, entonces β es una base para V y

$$[T]_\beta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Nuestro siguiente ejemplo es una aplicación de la diagonalización que será de interés en la Sección 5.3.

Ejemplo 15. Sea

$$A = \begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix}.$$

Demostraremos que A es diagonalizable y encontraremos una matriz Q de 2×2 tal que $Q^{-1}AQ$ sea una matriz diagonal. Esta información será luego utilizada para calcular A^n para cualquier entero positivo n.

Recuérdese que A es diagonalizable si y sólo si L_A es diagonalizable. Tenemos que el polinomio característico de L_A es $(t-1)(t-2)$; por tanto, L_A tiene dos eigenvalores diferentes y entonces L_A (y por tanto A) es diagonalizable. Para encontrar una base β para R^2 tal que $[L_A]_\beta$ sea una matriz diagonal, nótese que L_A tiene como eigenvalores a $\lambda_1 = 1$ y $\lambda_2 = 2$. Puede verse fácilmente que

$$\left\{ \begin{pmatrix} 2 \\ -1 \end{pmatrix} \right\}$$

es una base para E_{λ_1} y que

$$\left\{ \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$$

es una base para E_{λ_2}. Entonces para la base

$$\beta = \left\{ \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$$

tenemos

$$[L_A]_\beta = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Además, si

$$Q = \begin{pmatrix} 2 & 1 \\ -1 & -1 \end{pmatrix},$$

entonces, por el Teorema 5.1,

$$Q^{-1}AQ = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Finalmente, como

$$Q^{-1}AQ = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \qquad A = Q\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}Q^{-1}.$$

Entonces

$$\begin{aligned}
A^n &= \left[Q\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}Q^{-1}\right]^n \\
&= Q\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}Q^{-1}Q\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}Q^{-1} \cdots Q\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}Q^{-1} \\
&= Q\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}^n Q^{-1} \\
&= Q\begin{pmatrix} 1^n & 0 \\ 0 & 2^n \end{pmatrix}Q^{-1} \\
&= \begin{pmatrix} 2 & 1 \\ -1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 2^n \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & -2 \end{pmatrix} = \begin{pmatrix} 2-2^n & 2-2^{n+1} \\ -1+2^n & -1+2^{n+1} \end{pmatrix}.
\end{aligned}$$

Concluiremos esta sección con una aplicación que utilice la diagonalización para resolver un sistema de ecuaciones diferenciales.

Ejemplo 16. Considerar el sistema de ecuaciones diferenciales

$$\begin{cases} x_1' = 3x_1 + x_2 + x_3 \\ x_2' = 2x_1 + 4x_2 + 2x_3 \\ x_3' = -x_1 - x_2 + x_3, \end{cases}$$

donde, para toda i, $x_i = x_i(t)$ es una función diferenciable real, en la variable real t. Es evidente que este sistema tiene solución, que es la solución en la que cada $x_i(t)$ es la función cero. Determinaremos todas las soluciones del sistema.

Sea X: $R \to \mathsf{R}^3$ la función definida mediante

$$X(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix}.$$

La derivada de X se define como la función X', donde

$$X'(t) = \begin{pmatrix} x_1'(t) \\ x_2'(t) \\ x_3'(t) \end{pmatrix}.$$

Haciendo

$$A = \begin{pmatrix} 3 & 1 & 1 \\ 2 & 4 & 2 \\ -1 & -1 & 1 \end{pmatrix}$$

la matriz de coeficientes del sistema dado, podemos escribir este sistema en la forma matricial $X' = AX$, donde AX es el producto matricial de A por X.

El lector deberá verificar que A es diagonalizable y que si

$$Q = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ -1 & -1 & -1 \end{pmatrix},$$

entonces

$$Q^{-1}AQ = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

Hágase

$$D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix},$$

y sustitúyase $A = QDQ^{-1}$ en $X' = AX$ para encontrar $X' = QDQ^{-1}X$ o, de manera equivalente, $Q^{-1}X' = DQ^{-1}X$. Defínase a Y: $R \rightarrow \mathrm{R}^3$ mediante $Y(t) = Q^{-1}X(t)$. Puede demostrarse que Y es una función diferenciable y que, de hecho, $Y' = Q^{-1}X'$. Por lo tanto el sistema original puede escribirse como $Y' = DY$.

Como D es una matriz diagonal, el sistema $Y' = DY$ es fácil de resolver. Puesto que para

$$Y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{pmatrix},$$

entonces $Y' = DY$ puede escribirse como

$$\begin{pmatrix} y_1'(t) \\ y_2'(t) \\ y_3'(t) \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{pmatrix} = \begin{pmatrix} 2y_1(t) \\ 2y_2(t) \\ 4y_3(t) \end{pmatrix}.$$

Las tres ecuaciones

$$y_1'(t) = 2y_1(t)$$
$$y_2'(t) = 2y_2(t)$$
$$y_3'(t) = 4y_3(t)$$

son independientes entre sí y por tanto pueden resolverse individualmente. Se ve fácilmente (como en el Ejemplo 2 de la Sección 5.1) que la solución general de estas ecuaciones es $y_1(t) = c_1e^{2t}$, $y_2(t) = c_2e^{2t}$ y $y_3(t) = c_3e^{4t}$, donde c_1, c_2 y c_3 son escalares cualesquiera. Finalmente

$$\begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{pmatrix} = X(t) = QY(t) = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} c_1e^{2t} \\ c_2e^{2t} \\ c_3e^{4t} \end{pmatrix}$$
$$= \begin{pmatrix} c_1e^{2t} & & + \ c_3e^{4t} \\ & c_2e^{2t} & + \ 2c_3e^{4t} \\ -c_1e^{2t} & - \ c_2e^{2t} & - \ c_3e^{4t} \end{pmatrix}$$

da la solución general del sistema original. Nótese que esta solución puede escribirse como

$$X(t) = e^{2t}\left[c_1\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + c_2\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}\right] + e^{4t}\left[c_3\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}\right].$$

Las expresiones en los paréntesis rectangulares son sencillamente elementos cualesquiera de E_{λ_1} y E_{λ_2}, respectivamente, donde $\lambda_1 = 2$ y $\lambda_2 = 4$. Luego, la solución general del sistema original es $X(t) = e^{2t}z_1 + e^{4t}z_2$, donde $z_1 \in E_{\lambda_1}$ y $z_2 \in E_{\lambda_2}$.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.
   (a) Cualquier operador lineal en un espacio vectorial n-dimensional que tiene menos de n eigenvalores distintos no es diagonalizable.
   (b) Eigenvectores correspondientes al mismo eigenvalor son siempre linealmente dependientes.
   (c) Si un espacio vectorial es la suma directa de subespacios $W_1, W_2, \ldots, W_k$, entonces $W_i \cap W_j = \{0\}$ para $i \neq j$.
   (d) Si
   $$V = \sum_{i=1}^{k} W_i \quad \text{y} \quad W_i \cap W_j = \{0\} \qquad \text{para } i \neq j.$$
   entonces $V = W_1 \oplus W_2 \oplus \ldots \oplus W_k$.

(e) Si λ es un eigenvalor de un operador lineal T, entonces cada elemento de E_λ es un eigenvector de T.

(f) Si λ_1 y λ_2 son eigenvalores distintos de un operador lineal T, entonces $E_{\lambda_1} \cap E_{\lambda_2} = \{0\}$.

(g) Sean $A \in M_{n\times n}(F)$ y $\beta = \{x_1, \ldots, x_n\}$ una base para F^n formada de eigenvectores de A. Si Q es la matriz de $n \times n$ cuya columna i es $x_i (i = 1, 2, \ldots, n)$, entonces $Q^{-1}AQ$ es una matriz diagonal.

(h) Un operador lineal T en un espacio vectorial dimensionalmente finito es diagonalizable si y sólo si la multiplicidad de cada eigenvalor λ es igual a la dimensión de E_λ.

(i) Todo operador lineal diagonalizable tiene al menos un eigenvalor.

2. Para cada una de las matrices siguientes A en $M_{n\times n}(R)$, probar si A es diagonalizable y, en caso de serlo, encontrar una matriz Q tal que $Q^{-1}AQ$ sea una matriz diagonal.

(a) $\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ (b) $\begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$ (c) $\begin{pmatrix} 1 & 4 \\ 3 & 2 \end{pmatrix}$

(d) $\begin{pmatrix} 7 & -4 & 0 \\ 8 & -5 & 0 \\ 6 & -6 & 3 \end{pmatrix}$ (e) $\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix}$ (f) $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 3 \end{pmatrix}$

(g) $\begin{pmatrix} 3 & 1 & 1 \\ 2 & 4 & 2 \\ -1 & -1 & 1 \end{pmatrix}$

3. Para cada uno de los siguientes operadores lineales T, probar si T es diagonalizable y, en caso de serlo, encontrar una base β tal que $[T]_\beta$ sea una matriz diagonal.

(a) T: $P_3(R) \to P_3(R)$ definida mediante $T(f) = f' + f''$, donde f' y f'' son la primera y segunda derivadas de f, respectivamente.

(b) T: $P_2(R) \to P_2(R)$ definida mediante $T(ax^2 + bx + c) = cx^2 + bx + a$.

(c) T: $R^3 \to R^3$ definida mediante

$$T\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} a_2 \\ -a_1 \\ 2a_3 \end{pmatrix}$$

4. Demostrar la versión matricial del corolario al Teorema 5.10: Si $A \in M_{n\times n}(F)$ tiene n eigenvalores distintos, entonces A es diagonalizable.

5. Sean $W_1, W_2, \ldots, W_k$ subespacios de un espacio vectorial dimensionalmente finito V tal que

$$\sum_{i=1}^{k} W_i = V.$$

Demostrar que V es la suma directa de $W_1, W_2, \ldots, W_k$ si y sólo si

$$\dim(V) = \sum_{i=1}^{k} \dim(W_i).$$

6. Sea V un espacio vectorial dimensionalmente finito con una base $\beta = \{x_1, x_2, \ldots, x_n\}$, y sea $\beta_1, \beta_2, \ldots, \beta_k$ una partición de β (esto es, $\beta_1, \beta_2, \ldots, \beta_k$ son subconjuntos de β tales que $\beta = \beta_1 \cup \beta_2 \cup \ldots \cup \beta_k$ y $\beta_i \cap \beta_j = \varnothing$ si $i = j$). Demostrar que $V = L(\beta_1) \oplus L(\beta_2) \oplus \ldots \oplus L(\beta_k)$.

7. Enunciar y demostrar la versión matricial del Teorema 5.11.

8. (a) Justificar la prueba de diagonalizabilidad y el algoritmo para diagonalización enunciado en esta sección.
   (b) Enunciar el inciso (a) para matrices.

9. Si

$$A = \begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix} \in M_{2\times 2}(R),$$

encontrar A^n para cualquier entero positivo n.

10. Sea $A \in M_{n\times n}(F)$ tal que tenga dos eigenvalores distintos λ_1 y λ_2. Si $\dim(E_{\lambda_1}) = n - 1$, demostrar que A es diagonalizable.

11. Sea T un operador lineal en un espacio vectorial V dimensionalmente finito para el cual los distintos eigenvalores de T son $\lambda_1, \lambda_2, \ldots, \lambda_k$. Demostrar que

$$L(\{x \in V\colon\ x \text{ es un eigenvector de } T\}) = E_{\lambda_1} \oplus E_{\lambda_2} \oplus \ldots \oplus E_{\lambda_k}.$$

12. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V para el cual los distintos eigenvalores $\lambda_1, \lambda_2, \ldots, \lambda_k$ se presentan con multiplicidades $m_1, m_2, \ldots, m_k$, respectivamente. Si β es una base para V tal que $[T]_\beta$ es una matriz triangular, demostrar que los elementos de la diagonal de $[T]_\beta$ son $\lambda_1, \lambda_2, \ldots, \lambda_k$ y que cada λ_j aparece m_j veces ($j = 1, 2, \ldots, k$).

13. Supóngase que A es una matriz de $n \times n$ cuyo polinomio característico se descompone en un producto de factores de grado 1 y que los distintos eigenvalores de A son $\lambda_1, \lambda_2, \ldots \lambda_k$. Para cada j, sea m_j la multiplicidad de λ_j. Demostrar que

$$\operatorname{tr}(A) = \sum_{j=1}^{k} m_j\lambda_j.$$

14. Sea T un operador lineal invertible en un espacio vectorial dimensionalmente finito. Demostrar que T es diagonalizable si y sólo si T^{-1} es diagonalizable.

15. Sea $A \in M_{n \times n}(F)$. Demostrar que A es diagonalizable si y sólo si A^t es diagonalizable.

16. Encontrar la solución general del sistema de ecuaciones diferenciales.

$$\begin{cases} x_1' = 8x_1 + 10x_2 \\ x_2' = -5x_1 - 7x_2. \end{cases}$$

17. Sea

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

la matriz de coeficientes del sistema de ecuaciones diferenciales

$$\begin{cases} x_1' = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ x_2' = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdot \\ \cdot \\ \cdot \\ x_n' = a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n. \end{cases}$$

Supóngase que A es diagonalizable y que los distintos eigenvalores de A son $\lambda_1, \lambda_2, \ldots, \lambda_k$. Demostrar que una función diferenciable X: $R \to R^n$ es una solución del sistema si y sólo si X es de la forma

$$X(t) = e^{\lambda_1 t}z_1 + e^{\lambda_2 t}z_2 + \ldots + e^{\lambda_k t}z_k,$$

donde $z_i \in E_{\lambda_i}$ para $i = 1, 2, \ldots k$. Conclúyase que el conjunto de soluciones del sistema es un espacio vectorial real n-dimensional.

Los Ejercicios 18-20 se ocuparán del tema de la diagonalización simultánea.

Definiciones. *Dos operadores lineales* T *y* U *en el mismo espacio vectorial dimensionalmente finito* V *se denominan* simultáneamente diagonalizables *si existe una base* β *para* V *tal que* $[T]_\beta$ *y* $[U]_\beta$ *son ambas matrices diagonales. De la misma manera* A, $B \in M_{n \times n}(F)$ *se llaman* simultáneamente diagonalizables *si existe una matriz invertible* $Q \in M_{n \times n}(F)$ *tal que* $Q^{-1}AQ$ *y* $Q^{-1}BQ$ *son ambas matrices diagonales.*

18. (a) Si T y U son operadores lineales simultáneamente diagonalizables en un espacio vectorial dimensionalmente finito V, demostrar que $[T]_\beta$ y $[U]_\beta$ son matrices simultáneamente diagonalizables para cualquier base β.

(b) Demostrar que si A y B son matrices simultáneamente diagonalizables, entonces L_A y L_B son operadores simultáneamente diagonalizables.

19. (a) Demostrar que si T y U son operadores simultáneamente diagonalizables, entonces T y U conmutan: es decir, $TU = UT$.
(b) Demostrar que si A y B son matrices simultáneamente diagonalizables, entonces A y B conmutan.

Las recíprocas de (a) y (b) se establecerán en el Ejercicio 11 de la Sección 5.4.

20. Sea T un operador lineal diagonalizable en un espacio vectorial dimensionalmente finito y sea m cualquier entero positivo. Demostrar que T y T^m son simultáneamente diagonalizables.

21. Sean $W_1, W_2, K_1, K_2, \ldots, K_p, M_1, M_2, \ldots, M_q$ subespacios de un espacio vectorial V tales que $W_1 = K_1 \oplus K_2 \oplus \ldots \oplus K_p$ y $W_2 = M_1 \oplus M_2 \oplus \ldots \oplus M_q$. Demostrar que si $W_1 \cap W_2 = \{0\}$, entonces $W_1 + W_2 = W_1 \oplus W_2 = K_1 \oplus K_2 \oplus \ldots \oplus K_p \oplus M_1 \oplus M_2 \oplus \ldots \oplus M_q$.

5.3* LIMITES DE MATRICES Y CADENAS DE MARKOV

Si A es una matriz cuadrada con elementos complejos, entonces, para cualquier entero positivo m, A^m es una matriz cuadrada del mismo tamaño y que también tiene elementos complejos. En muchas de las ciencias naturales y de la vida existen aplicaciones prácticas de importancia que requieren de la determinación del "límite" (si existe alguno) de la secuencia de matrices $A, A^2, A^3, \ldots$ En esta sección consideraremos tales límites y examinaremos una situación importante en la que surge esta clase de límites.

Definiciones. *Sean* $L, A_1, A_2, A_3, \ldots$ *matrices de* $n \times p$ *con elementos complejos. Se dice que la sucesión* $A_1, A_2, A_3, \ldots$ converge *a la matriz* L, *denominada el* límite *de la sucesión, si para cada* i $(1 \leq i \leq n)$ *y* $j (1 \leq j \leq p)$ *la sucesión de números complejos* $(A_1)_{ij}, (A_2)_{ij}, (A_3)_{ij} \ldots$ *converge a* L_{ij}. *(El límite de la sucesión de números complejos* $\{z_m: m = 1, 2, \ldots\}$, *donde* $z_m = r_m + is_m$ *siendo* r_m *y* s_m *números reales, queda definido en términos de los límites de la sucesión de las partes real e imaginaria como*

$$\lim_{m\to\infty} z_m = (\lim_{m\to\infty} r_m) + i(\lim_{m\to\infty} s_m).)$$

Para expresar el hecho de que la sucesión $A_1, A_2, A_3, \ldots$ *converge a* L, *debemos escribir* $\lim_{m\to\infty} A_m = L$.

Ejemplo 17. Si

$$A_m = \begin{pmatrix} 1 - \frac{1}{m} & \left(-\frac{3}{4}\right)^m & \frac{3m^2}{m^2+1} + i\left(\frac{2m+1}{m-1}\right) \\ \left(\frac{i}{2}\right)^m & 2 & \left(1 + \frac{1}{m}\right)^m \end{pmatrix},$$

entonces

$$\lim_{m\to\infty} A_m = \begin{pmatrix} 1 & 0 & 3+2i \\ 0 & 2 & e \end{pmatrix},$$

donde e es la base de los logaritmos naturales.

Una sencilla pero muy importante propiedad de los límites de las matrices está contenida en el teorema siguiente. Nótese la analogía con la propiedad ordinaria de límites de sucesiones de números reales que asegura que si $\lim_{m\to\infty} a_m$ existe, entonces

$$\lim_{m\to\infty} ca_m = c(\lim_{m\to\infty} a_m).$$

Teorema 5.15. *Sea* $A_1, A_2, A_3, \ldots$ *una sucesión de matrices de* $n \times p$ *con elementos complejos tales que*

$$\lim_{m\to\infty} A_m = L \in M_{n\times p}(C).$$

Entonces para cualquier $B \in M_{r\times n}(C)$ *y* $C \in M_{p\times s}(C)$,

$$\lim_{m\to\infty} BA_m = BL \quad \text{y} \quad \lim_{m\to\infty} A_mC = LC.$$

DEMOSTRACIÓN. Para cualquier $i(1 \leq i \leq r)$ y $j(1 \leq j \leq p)$,

$$\lim_{m\to\infty} [(BA_m)_{ij}] = \lim_{m\to\infty} \left[\sum_{k=1}^{n} B_{ik}(A_m)_{kj}\right]$$

$$= \sum_{k=1}^{n} B_{ki}\{\lim_{m\to\infty} [(A_m)_{kj}]\} = \sum_{k=1}^{n} B_{ik}L_{kj} = (BL)_{ij}.$$

Por lo tanto $\lim_{m\to\infty} BA_m = BL$. La demostración de que $\lim_{m\to\infty} A_mC = LC$ es semejante. ■

Corolario. *Sea* $A \in M_{n\times n}(C)$ *y sea* $\lim_{m\to\infty} A^m = L$. *Entonces para cualquier matriz invertible* $Q \in M_{n\times n}(C)$,

$$\lim_{m\to\infty} (QAQ^{-1})^m = QLQ^{-1}.$$

DEMOSTRACIÓN. Puesto que

$$(QAQ^{-1})^m = (QAQ^{-1})(QAQ^{-1})\ldots(QAQ^{-1}) = QA^mQ^{-1},$$

tenemos, de acuerdo con el Teorema 5.15, que

$$\lim_{m\to\infty} [(QAQ^{-1})^m] = \lim_{m\to\infty} (QA^mQ^{-1}) = Q(\lim_{m\to\infty} A^m)Q^{-1} = QLQ^{-1}. \quad \blacksquare$$

El siguiente resultado importante proporciona las condiciones necesarias y suficientes para la existencia de la clase de límite que estamos considerando.

Teorema 5.16. *Sea* A *una matriz cuadrada con elementos complejos. Entonces* $\lim_{m\to\infty}$ A^m *existe si y sólo si se satisfacen las condiciones siguientes.*

(a) *Si* λ *es un eigenvalor de* A, *entonces* $|\lambda| \leq 1$.
(b) *Si* λ *es un eigenvalor de* A *tal que* $|\lambda| = 1$, *entonces* λ *es el número real* 1.
(c) *Si* 1 *es un eigenvalor de* A, *entonces la dimensión del eigenespacio correspondiente a* 1 *es igual a la multiplicidad de* 1 *como eigenvalor de* A.

Desafortunadamente no será posible demostrar la suficiencia de estas condiciones ni la necesidad de la condición (c) hasta que estudiemos la forma canónica de Jordan. Por esta razón la demostración del teorema será pospuesta hasta la Sección 6.2 (Ejercicio 18). Sin embargo, la necesidad de las dos primeras condiciones se infiere fácilmente del hecho de que $\lim_{m\to\infty} \lambda^m$ si y sólo si $\lambda = 1$ o bien $|\lambda| < 1$. (Puede demostrarse que este caso, que sin duda el lector conoce para los números reales λ, también se cumple para los complejos.) Supóngase entonces que λ es un eigenvalor de A para el que las condiciones (a) y (b) fallan, esto es, tal que $|\lambda| > 1$ o bien que $|\lambda| = 1$ pero $\lambda \neq 1$. Sea x un eigenvector de A que corresponda a λ. Considerando a x como una matriz de $n \times 1$ vemos que, de acuerdo con el Teorema 5.15,

$$\lim_{m\to\infty} (A^m x) = (\lim_{m\to\infty} A^m)x = Lx,$$

donde $L = \lim_{m\to\infty} A^m$. Pero $\lim_{m\to\infty} (A^m x) = \lim_{m\to\infty} (\lambda^m x)$ es divergente puesto que $\lim_{m\to\infty} \lambda^m$ no existe. Por lo tanto si $\lim_{m\to\infty} A^m$ existe, entonces las condiciones (a) y (b) del Teorema 5.16 deben satisfacerse. Aun cuando en este momento no seamos capaces de demostrar la necesidad de la tercera condición, consideremos un ejemplo en el que esta condición falle. Obsérvese que para la matriz

$$B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

el eigenvalor $\lambda = 1$ tiene multiplicidad 2, mientras que $\dim(E_\lambda) = 1$. Pero por inducción simple

$$B^m = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$$

y por lo tanto $\lim_{m\to\infty} B^m$ no existe. (Veremos más adelante que si A es una matriz para la que la condición (c) falla, entonces puede escogerse la forma canónica de Jordan de A de forma tal que su submatriz izquierda superior de 2×2 sea justamente esta matriz B.)

Sin embargo, en la mayor parte de las aplicaciones que involucran este tipo de límite, la matriz A es diagonalizable. Cuando la condición (c) del Teorema 5.16 se sustituye por la condición de mayor fuerza de que A es diagonalizable (ver el Teorema 5.14), entonces se puede demostrar fácilmente la existencia del límite.

Teorema 5.17. *Sea* $A \in M_{n\times n}(C)$ *tal que las condiciones siguientes se satisfacen:*

(a) *Si* λ *es un eigenvalor de* A, *entonces* $|\lambda| \leq 1$.
(b) *Si* λ *es un eigenvalor de* A *tal que* $|\lambda| = 1$, *entonces* λ *es el número real* 1.
(c) A *es diagonalizable.*

Entonces existe $\lim_{m\to\infty} A^m$.

DEMOSTRACIÓN. Como A es diagonalizable, existe una matriz invertible Q tal que $Q^{-1}AQ = D$ es una matriz diagonal. Sea

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Dado que $\lambda_1, \lambda_2, \ldots, \lambda_n$ son los eigenvalores de A, las condiciones (a) y (b) muestran que $\lambda_i = 1$ o bien $|\lambda_i| < 1$ para $1 \leq i \leq n$. Por lo tanto

$$\lim_{m\to\infty} \lambda_i^m = \begin{cases} 1 & \text{si } \lambda_i = 1 \\ 0 & \text{en cualquier otro caso.} \end{cases}$$

Pero como

$$D^m = \begin{pmatrix} \lambda_1^m & 0 & \cdots & 0 \\ 0 & \lambda_2^m & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & \lambda_n^m \end{pmatrix},$$

la sucesión $D, D^2, D^3, \ldots$ converge a un límite L. Por lo tanto, por el corolario del Teorema 5.15,

$$\lim_{m\to\infty} A^m = \lim_{m\to\infty} (QDQ^{-1})^m = QLQ^{-1}. \quad \blacksquare$$

La técnica para calcular $\lim_{m\to\infty} A^m$, utilizada en la demostración del teorema anterior, es de gran utilidad. Utilizaremos este método para calcular $\lim_{m\to\infty} A^m$ para la matriz

$$A = \begin{pmatrix} \frac{7}{4} & -\frac{9}{4} & -\frac{15}{4} \\ \frac{3}{4} & \frac{7}{4} & \frac{3}{4} \\ \frac{3}{4} & -\frac{9}{4} & -\frac{11}{4} \end{pmatrix}.$$

Si

$$Q = \begin{pmatrix} 1 & 3 & -1 \\ -3 & -2 & 1 \\ 2 & 3 & -1 \end{pmatrix},$$

entonces

$$D = Q^{-1}(AQ) = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 1 & 2 \\ -5 & 3 & 7 \end{pmatrix} \begin{pmatrix} 1 & -\frac{3}{2} & -\frac{1}{4} \\ -3 & 1 & \frac{1}{4} \\ 2 & -\frac{3}{2} & -\frac{1}{4} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{4} \end{pmatrix}.$$

Por lo tanto

$$\lim_{m\to\infty} A^m = \lim_{m\to\infty} (QDQ^{-1})^m = \lim_{m\to\infty} (QD^mQ^{-1}) = Q(\lim_{m\to\infty} D^m)Q^{-1}$$

$$= \begin{pmatrix} 1 & 3 & -1 \\ -3 & -2 & 1 \\ 2 & 3 & -1 \end{pmatrix} \left[\lim_{m\to\infty} \begin{pmatrix} 1 & 0 & 0 \\ 0 & (-\frac{1}{2})^m & 0 \\ 0 & 0 & (\frac{1}{4})^m \end{pmatrix} \right] \begin{pmatrix} -1 & 0 & 1 \\ -1 & 1 & 2 \\ -5 & 3 & 7 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 3 & -1 \\ -3 & -2 & 1 \\ 2 & 3 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} -1 & 0 & 1 \\ -1 & 1 & 2 \\ -5 & 3 & 7 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 1 \\ 3 & 0 & -3 \\ -2 & 0 & 2 \end{pmatrix}.$$

Consideremos ahora un ejemplo sencillo en el que se presenta el límite de las potencias de una matriz. Supóngase que la población de cierta área metropolitana se mantiene constante pero que hay un movimiento continuo de gente entre la ciudad y los suburbios. Específicamente, sean los elementos de la siguiente matriz A las posibilidades de que alguien que vive en la ciudad o en los suburbios el primero de enero estará viviendo en cada región el primero de enero del año siguiente.

	Viven actualmente en la ciudad	*Viven actualmente en los suburbios*	
Vivirán el próximo año en la ciudad	0.90	0.02	$= A$
Vivirán el próximo año en los suburbios	0.10	0.98	

Entonces, por ejemplo, la probabilidad de que alguien que vive en la ciudad (el primero de enero) estará viviendo en los suburbios el próximo año (el primero de enero) es 0.10. Nótese que como los elementos de cada columna de A representan las probabilidades de residencia en cada uno de los dos sitios, los elementos de A son no negativos. Más aún, la suposición de una población constante en el área metropolitana requiere que la suma de los elementos de cada columna de A sea 1. Cualquier matriz que tenga estas dos propiedades (que los elementos sean no negativos y que la suma de los elementos de cada columna sea 1) se llama *matriz de transición* (o *matriz estocástica).* Para una matriz de transición M de $n \times n$ arbitraria, los renglones y las columnas corresponden a n *estados* y el elemento M_{ij} representa la probabilidad de pasar del estado j al estado i en una *etapa.* En nuestro ejemplo, se tienen dos estados (residir en la ciudad y residir en los suburbios), y A_{21} representa la probabilidad de emigrar de la ciudad a los suburbios en una etapa (año).

Determinemos ahora la probabilidad de que un residente de la ciudad esté residiendo en los suburbios después de dos años. Obsérvese primero que hay dos maneras diferentes en las que tal cambio puede haberse realizado —ya sea permaneciendo en la ciudad un año y después mudándose a los suburbios, o mudándose a los suburbios durante el primer año y permaneciendo en ellos en el segundo (véase Fig. 5.3). La probabilidad de que un habitante de la ciudad permanezca en la ciudad el siguiente año es 0.90 y la probabilidad de que un habitante de la ciudad se mude a los suburbios durante el próximo año es de 0.10. Por lo tanto, la probabilidad de que un residente de la ciudad permanezca en la ciudad durante un año y se mude a los suburbios durante el siguiente es 0.90 (0.10). De la misma manera, la probabilidad de que un habitante de la ciudad se mude a los suburbios durante el primer año y permanezca ahí durante el siguiente es 0.10(0.98). De este modo la probabilidad de que un residente de la ciudad esté viviendo en los suburbios después de 2 años es $0.90(0.10) + 0.10(0.98) = 0.188$. Obsérvese que este número se obtuvo

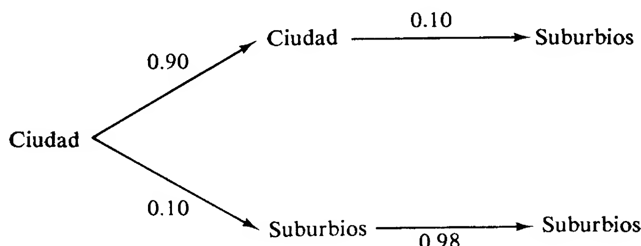


figura 5.3

mediante la misma operación que la que produce $(A^2)_{21}$; por lo tanto $(A^2)_{21}$ representa la probabilidad de que un habitante de la ciudad esté residiendo en los suburbios después de dos años. En general, para cualquier matriz de transición M $(M^m)_{ij}$ representa la probabilidad de pasar del estado j al estado i en m etapas.

Supóngase además que 70% de la población de 1970 del área metropolitana vivía en la ciudad y 30% vivía en los suburbios. Registremos esta información como un vector columna:

$$\begin{matrix}\text{Proporción de habitantes de la ciudad} \\ \text{Proporción de residentes de los suburbios}\end{matrix} \begin{pmatrix} 0.70 \\ 0.30 \end{pmatrix} = P.$$

Nótese que los renglones de P corresponden a los estados de residir en la ciudad y de residir en los suburbios, respectivamente —el mismo orden en que los estados están registrados en la matriz de transición A. Obsérvese también que P es un vector columna que contiene elementos no negativos cuya suma es 1; tal vector se denomina *vector de probabilidad.* En esta terminología cada columna de una matriz de transición es un vector de probabilidad.

Consideremos ahora el significado del vector AP. La primera coordenada de este vector está formada por la operación $0.90(0.70) + 0.02(0.30)$. El término $0.90(0.70)$ representa la proporción de la población metropolitana de 1970 que permaneció en la ciudad durante el siguiente año, y el término $0.02(0.30)$ representa la proporción de la población metropolitana de 1970 que se mudó a la ciudad durante el año siguiente. Por lo tanto, la primera coordenada de AP representa la proporción de la población metropolitana que vivía en la ciudad 1 año después de 1970. De la misma manera la segunda coordenada de

$$AP = \begin{pmatrix} 0.636 \\ 0.364 \end{pmatrix}$$

representa la proporción de la población metropolitana que vivía en los suburbios en 1971. Este argumento puede extenderse fácilmente para demostrar que las coordenadas de

$$A^2P = A(AP) = \begin{pmatrix} 0.57968 \\ 0.42032 \end{pmatrix}$$

representan las proporciones de la población metropolitana que estaban viviendo en cada uno de los sitios en 1972. En general, las coordenadas de A^mP representan la proporción de la población metropolitana que vivirá en la ciudad y en los suburbios, respectivamente, después de m etapas (m años después de 1970).

¿Si esta tendencia continúa se vaciará la ciudad? En vista de lo antes expuesto es natural definir la proporción eventual de habitantes de la ciudad y de los suburbios como la primera y segunda coordenadas, res-

pectivamente, de $\lim_{m\to\infty} A^m P$. Calculemos este límite. Utilizando la notación anterior

$$D = Q^{-1}AQ = \begin{pmatrix} \frac{1}{6} & \frac{1}{6} \\ \frac{5}{6} & -\frac{1}{6} \end{pmatrix} \begin{pmatrix} 0.90 & 0.02 \\ 0.10 & 0.98 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 5 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0.88 \end{pmatrix};$$

luego entonces

$$L = \lim_{m\to\infty} A^m = \lim_{m\to\infty} (QD^mQ^{-1}) = Q\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} Q^{-1} = \begin{pmatrix} \frac{1}{6} & \frac{1}{6} \\ \frac{5}{6} & \frac{5}{6} \end{pmatrix}.$$

Por lo tanto

$$\lim_{m\to\infty} A^m P = LP = \begin{pmatrix} \frac{1}{6} \\ \frac{5}{6} \end{pmatrix};$$

de manera que $\frac{1}{6}$ de la población vivirá en la ciudad y $\frac{5}{6}$ de la población vivirá en los suburbios. Es fácil demostrar que en este ejemplo

$$LP = \begin{pmatrix} \frac{1}{6} \\ \frac{5}{6} \end{pmatrix}$$

para cualquier vector de probabilidad P. ¡Por lo tanto en este ejemplo las proporciones eventuales de habitantes de la ciudad y de los suburbios son independientes de las proporciones iniciales (dadas por el vector P)!

Al analizar el problema ciudad-suburbios dimos interpretaciones probabilistas de A^2 y de AP, demostrando que A^2 es una matriz de transición y AP es un vector de probabilidad. Se pueden utilizar argumentos semejantes para demostrar que el producto de dos matrices de transición es una matriz de transición y que el producto de una matriz de transición por un vector de probabilidad es un vector de probabilidad. Una demostración alternativa de estos resultados puede basarse en el teorema siguiente, que caracteriza a las matrices de transición y a los vectores de probabilidad.

Teorema 5.18. *Sea* M *una matriz de* $n \times n$ *con elementos (reales) no negativos, sea* x *un vector columna en* R^n *de coordenadas no negativas y sea* $u \in \mathsf{R}^n$ *el vector columna en el que todas las coordenadas son iguales a* 1. *Entonces:*

(a) M *es una matriz de transición si y sólo si* $M^t u = u$.
(b) x *es un vector de probabilidad si y sólo si* $u^t x = (1)$.

DEMOSTRACIÓN. Ejercicio.

Corolario.

(a) *El producto de dos matrices de transición de* $n \times n$ *es una matriz de transición de* $n \times n$. *En particular, cualquier potencia de una matriz de transición es una matriz de transición.*

(b) *El producto de una matriz de transición por un vector de probabilidad es un vector de probabilidad.*

DEMOSTRACIÓN. Ejercicio.

Un *proceso estocástico* tiene como finalidad predecir el estado de un objeto que esté restringido a estar exactamente en uno de ciertos posibles estados en un instante dado cualquiera, pero que cambia de estado de alguna manera aleatoria. Normalmente, la probabilidad de que el objeto se encuentre en un estado particular en un instante dado dependerá de factores tales como:

1. El estado en cuestión.
2. El instante en cuestión.
3. Algunos o todos los estados anteriores en los cuales estuvo el objeto.
4. Los estados en los que se encuentran otros objetos o en los que se hayan encontrado.

Por ejemplo, el objeto podría ser un votante americano y el estado del objeto podría ser su preferencia por algún partido político, o el objeto podría ser una molécula de H_2O y los estados podrían ser los estados físicos en los cuales el H_2O puede existir (los estados sólido, líquido y gaseoso). En estos ejemplos los cuatro factores antes mencionados influenciarán la probabilidad de que los objetos se encuentren en un estado particular en un instante particular.

Sin embargo, si la probabilidad de que un objeto que está en un estado cambie a otro estado diferente depende únicamente de los dos estados (y no del tiempo, estados anteriores u otros factores), entonces el proceso estocástico se denomina *proceso de Markov*. Además, si el número de estados posibles es finito, entonces el proceso de Markov se llama *cadena de Markov*. El ejemplo anterior del movimiento de población entre la ciudad y los suburbios es una cadena de Markov de dos estados.

Consideremos otra cadena de Markov. Un cierto plantel de bachillerato desearía obtener información sobre la probabilidad de que se gradúen las distintas clases de estudiantes actualmente inscritos. La escuela clasifica a un estudiante como de segundo o de primer grado dependiendo del número de créditos que el estudiante haya contabilizado. Los datos con que cuenta la escuela indican que de un semestre de otoño al siguiente se graduará el 40% de los estudiantes de segundo año, el 30% continuará en el mismo grado, y el 30% abandonará los estudios definitivamente. Para los de primer año, los datos muestran que el 10% se graduará para el próximo otoño, 50% pasará al segundo grado, 20% permanecerá en el mismo grado y 20% abandonará definitivamente la carrera. En este año el 50% de los estudiantes de la escuela son de segundo grado y el 50%

de primer grado. Suponiendo que la tendencia indicada por los datos se prolonga indefinidamente, la escuela desearía saber

1. El porcentaje de los actuales estudiantes que se graduará, el porcentaje de los que pasarán a segundo grado, el porcentaje de los que estarán en el primer grado y el porcentaje de los que abandonarán definitivamente para el próximo otoño.
2. Los mismos porcentajes que en el inciso 1 para el semestre de otoño de aquí a 2 años.
3. El porcentaje de sus estudiantes actuales que eventualmente se graduará.

El párrafo anterior describe a una cadena de Markov de cuatro estados que son:

1. Haberse graduado.
2. Estar en el segundo grado.
3. Estar en el primer grado.
4. Haber abandonado definitivamente.

Los datos antes mencionados nos proporcionan la matriz de transición

$$A = \begin{pmatrix} 1 & 0.4 & 0.1 & 0 \\ 0 & 0.3 & 0.5 & 0 \\ 0 & 0 & 0.2 & 0 \\ 0 & 0.3 & 0.2 & 1 \end{pmatrix}$$

de la cadena de Markov. (Nótese que los estudiantes ya graduados o que han abandonado de una manera definitiva se considera que permanecen de una manera indefinida en sus estados respectivos, por lo que un estudiante del primer grado que abandona la escuela y que regresa semestres después, no se considera que haya cambiado de estado, se supone que el estudiante permaneció en el estado de ser de primer grado durante el tiempo que no se inscribió.) Además, se nos informa que la distribución actual de los estudiantes es tal que la mitad de ellos se encuentra respectivamente en los estados 2 y 3 y ninguno en los estados 1 y 4. El vector

$$P = \begin{pmatrix} 0 \\ 0.5 \\ 0.5 \\ 0 \end{pmatrix}$$

que describe la probabilidad inicial de estar en cualquier estado se llama *vector de probabilidad inicial* para la cadena de Markov.

Con el objeto de resolver la primera pregunta debemos determinar la probabilidad de que un estudiante actual esté en cualquiera de los estados para el próximo otoño. Como se vio anteriormente, estas probabilidades están dadas por las coordenadas del vector

$$AP = \begin{pmatrix} 0.25 \\ 0.40 \\ 0.10 \\ 0.25 \end{pmatrix}.$$

Por lo tanto, para el próximo otoño, el 25% de los estudiantes actuales se graduará, el 40% estará en el segundo grado, el 10% estará en el primer grado y el 25% abandonará la escuela. De la misma manera

$$A^2P = A(AP) = \begin{pmatrix} 0.42 \\ 0.17 \\ 0.02 \\ 0.39 \end{pmatrix}$$

proporciona la información requerida para resolver la pregunta 2: dentro de dos años el 42% de los estudiantes actuales se graduará, el 17% estará en el segundo grado, 2% estará en el primer grado y el 39% abandonará la escuela.

Finalmente, la respuesta a la pregunta 3 la proporciona el vector LP, donde $L = \lim_{m\to\infty} A^m$. El lector deberá verificar que si

$$Q = \begin{pmatrix} 1 & -4 & 19 & 0 \\ 0 & 7 & -40 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & -3 & 13 & 1 \end{pmatrix},$$

entonces

$$D = Q^{-1}AQ = \begin{pmatrix} 1 & \frac{4}{7} & \frac{27}{56} & 0 \\ 0 & \frac{1}{7} & \frac{5}{7} & 0 \\ 0 & 0 & \frac{1}{8} & 0 \\ 0 & \frac{3}{7} & \frac{29}{56} & 1 \end{pmatrix} \begin{pmatrix} 1 & 0.4 & 0.1 & 0 \\ 0 & 0.3 & 0.5 & 0 \\ 0 & 0 & 0.2 & 0 \\ 0 & 0.3 & 0.2 & 1 \end{pmatrix} \begin{pmatrix} 1 & -4 & 19 & 0 \\ 0 & 7 & -40 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & -3 & 13 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0 \\ 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

De donde

$$L = \lim_{m\to\infty} A^m = Q(\lim_{m\to\infty} D^m)Q^{-1}$$

$$= \begin{pmatrix} 1 & -4 & 19 & 0 \\ 0 & 7 & -40 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & -3 & 13 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{4}{7} & \frac{27}{56} & 0 \\ 0 & \frac{1}{7} & \frac{5}{7} & 0 \\ 0 & 0 & \frac{1}{8} & 0 \\ 0 & \frac{3}{7} & \frac{29}{56} & 1 \end{pmatrix} = \begin{pmatrix} 1 & \frac{4}{7} & \frac{27}{56} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{3}{7} & \frac{29}{56} & 1 \end{pmatrix}.$$

De este modo

$$LP = \begin{pmatrix} \frac{59}{112} \\ 0 \\ 0 \\ \frac{53}{112} \end{pmatrix},$$

y por lo tanto la probabilidad de que uno de los actuales estudiantes se gradúe es de $^{59}/_{112}$.

En los dos ejemplos anteriores hemos visto que $\lim_{m\to\infty} A^mP$, donde A es la matriz de transición y P es el vector de probabilidad inicial de la cadena de Markov, permite conocer las proporciones eventuales en cada estado. En general, sin embargo, no es necesario que exista el límite de las potencias de una matriz de transición. Por ejemplo, si

$$M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

es evidente que $\lim_{m\to\infty} M^m$ no existe. (Potencias impares de M son iguales a M, y potencias pares de M son iguales a I.) La razón por la cual el límite no existe, es que la condición (b) del Teorema 5.16 no se satisface para M (-1 es un eigenvalor). De hecho, puede demostrarse (véase Ejercicio 20 de la Sección 6.2) que las únicas matrices de transición A tales que $\lim_{m\to\infty} A^m$ no existe son precisamente aquellas matrices en las que la condición (b) del Teorema 5.16 no se cumple.

Pero aun cuando exista el límite de las potencias de la matriz de transición, el cálculo de dicho límite puede llegar a ser muy difícil. (Se sugiere al lector realizar el Ejercicio 6 para comprobar la verdad de la última oración.) Afortunadamente, existe una clase de matrices de transición grande e importante para las cuales el límite existe y es fácil de calcular —esta es la clase de matrices "regulares" de transición.

Definición. *Si alguna potencia de una matriz de transición contiene únicamente elementos positivos, entonces la matriz se llama* matriz regular de transición.

Ejemplo 18. La matriz de transición

$$\begin{pmatrix} 0.90 & 0.02 \\ 0.10 & 0.98 \end{pmatrix}$$

de la cadena de Markov que describe el movimiento de población entre la ciudad y los suburbios es claramente regular puesto que todos los elementos son positivos. Por otra parte, la matriz de transición

$$A = \begin{pmatrix} 1 & 0.4 & 0.1 & 0 \\ 0 & 0.3 & 0.5 & 0 \\ 0 & 0 & 0.2 & 0 \\ 0 & 0.3 & 0.2 & 1 \end{pmatrix}$$

de la cadena de Markov que describe las inscripciones del plantel de bachillerato no es regular. (Es fácil demostrar que la primera columna de A^m es

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

para toda m; de donde $(A^m)_{41}$, por ejemplo, no es nunca positiva.)

Obsérvese que una matriz regular de transición puede contener elementos nulos; por ejemplo,

$$M = \begin{pmatrix} 0.9 & 0.5 & 0 \\ 0 & 0.5 & 0.4 \\ 0.1 & 0 & 0.6 \end{pmatrix}$$

es regular puesto que todos los elementos de M^2 son positivos.

En el resto de esta sección nos dedicaremos principalmente a demostrar que si A es una matriz regular de transición, entonces existe $L = \lim_{m \to \infty} A^m$ y las columnas de L son idénticas. (Recuérdese la forma de L en el problema ciudad-suburbio.) Con este hecho será fácil calcular el límite. Durante el transcurso de la demostración de este resultado obtendremos algunos teoremas interesantes sobre la magnitud de los eigenvalores de cualquier matriz cuadrada. Estas cotas estarán dadas en términos de la suma de los valores absolutos de los elementos de los renglones y de las columnas de la matriz. La terminología necesaria se introduce en la definición siguiente.

Definiciones. *Sea* $A \in M_{n \times n}(C)$. *Defínase a* $\rho_i(A)$ *como la suma de los valores absolutos de los elementos del renglón* i *de* A *y* $\nu_j(A)$ *como la suma de los valores absolutos de los elementos de la columna* j *de* A. *Entonces*

$$\rho_i(A) = \sum_{j=1}^{n} |A_{ij}| \qquad \textit{para } i = 1, 2, \ldots, n$$

y

$$\nu_j(A) = \sum_{j=1}^{n} |A_{ij}| \qquad \textit{para } j = 1, 2, \ldots, n.$$

La suma de renglones *de* A, *denotada por* $\rho(A)$, *y la* suma de columnas *de* A *denotada por* $\nu(A)$, *se definen como*

$$\rho(A) = \max \{\rho_i(A): \ 1 \leq i \leq n\} \ \ y \ \ \nu(A) = \max \{\nu_j(A): \ 1 \leq j \leq n\}.$$

Ejemplo 19. Para la matriz

$$A = \begin{pmatrix} 1 & -1 & 5 \\ -4 & 0 & 6 \\ 3 & 2 & -1 \end{pmatrix},$$

$\rho_1(A) = 7$, $\rho_2(A) = 10$, $\rho_3(A) = 6$, $\nu_1(A) = 8$, $\nu_2(A) = 3$ y $\nu_3(A) =$ $= 12$. Por lo tanto $\rho(A) = 10$ y $\nu(A) = 12$.

Nuestros siguientes resultados muestran que el menor de entre $\rho(A)$ y $\nu(A)$ es una cota superior para el valor absoluto de los eigenvalores de A. En el ejemplo anterior, por ejemplo, A no tiene ningún eigenvalor cuyo valor absoluto sea mayor de 10.

Teorema 5.19. *Sea* λ *un eigenvalor de* $A \in M_{n\times n}(C)$. *Entonces* $|\lambda| \leq \rho(A)$.

DEMOSTRACIÓN. Sea

$$x = \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix}$$

un eigenvector de A para el que λ es el eigenvalor correspondiente. Entonces x satisface la ecuación matricial $Ax = \lambda x$ que puede ser escrita como el sistema de ecuaciones lineales

$$\sum_{j=1}^{n} A_{ij}x_j = \lambda x_i \qquad (i = 1, 2, \ldots, n). \tag{4}$$

Supóngase que x_k es la coordenada de x que tiene el mayor valor absoluto y sea $b = |x_k|$. De la k-ésima ecuación de (4) tenemos

$$\begin{aligned} |\lambda| b = |\lambda||x_k| = |\lambda x_k| &= \Big|\sum_{j=1}^{n} A_{kj}x_j\Big| \leq \sum_{j=1}^{n} |A_{kj}x_j| \\ &= \sum_{j=1}^{n} |A_{kj}||x_j| \leq \sum_{j=1}^{n} |A_{kj}| b = \rho_k(A)b \leq \rho(A)b. \end{aligned} \tag{5}$$

Pero como x que es un eigenvector no nulo, $b \neq 0$. Luego, dividiendo ambos lados de la ecuación (5) por b, se obtiene $|\lambda| \leq \rho(A)$. ■

Corolario 1. *Sea λ un eigenvalor de* $A \in M_{n \times n}(C)$. *Entonces* $|\lambda| \leq min \{\rho(A), \nu(A)\}$.

DEMOSTRACIÓN. Como por el Teorema 5.19 $|\lambda| \leq \rho(A)$, es suficiente demostrar que $|\lambda| \leq \nu(A)$.

El Ejercicio 14 de la Sección 5.1 muestra que λ es un eigenvalor de A^t. Por lo tanto, de acuerdo con el Teorema 5.19 $|\lambda| \leq \rho(A^t)$. Pero los renglones de A^t son las columnas de A. Por lo tanto $\rho(A^t) = \nu(A)$, y así $|\lambda| \leq \nu(A)$. ■

Como la suma de los elementos de las columnas de una matriz de transición es 1, se obtiene de inmediato la siguiente conclusión a partir del Corolario 1.

Corolario 2. *Si λ es un eigenvalor de una matriz de transición, entonces* $|\lambda| \leq 1$.

El siguiente resultado muestra que se ha alcanzado la cota superior en el corolario anterior.

Teorema 5.20. *Toda matriz de transición tiene a* 1 *como eigenvalor.*

DEMOSTRACIÓN. Sea A una matriz de transición de $n \times n$ y sea $u \in R^n$ el vector columna en donde cada coordenada es 1. Entonces, según el Teorema 5.18, $A^t u = u$ y por tanto u es un eigenvector de A^t que corresponde al eigenvalor 1. Pero como A y A^t tienen los mismos eigenvalores se tiene que 1 también es un eigenvalor de A. ■

Supóngase ahora que A es una matriz de transición para la que algún eigenvector que corresponde al eigenvalor 1 tiene únicamente coordenadas no negativas. Entonces algún múltiplo de este vector será un vector de probabilidad P, así como un eigenvector de A que corresponde al eigenvalor 1. Es interesante observar que si P es el vector de probabilidad inicial de una cadena de Markov que tenga a A como matriz de transición, entonces la cadena de Markov es completamente estática, pues en esta situación $A^m P = P$ para cualquier entero positivo m y por lo tanto la probabilidad de estar en cada estado nunca cambia. Considérese, por ejemplo, el problema ciudad-suburbios con

$$P = \begin{pmatrix} \frac{1}{6} \\ \frac{5}{6} \end{pmatrix}.$$

Teorema 5.21. *Sea* $A \in M_{n \times n}(C)$ *una matriz en la que todos los elementos son positivos y sea λ un eigenvalor de* A *tal que* $|\lambda| = \rho(A)$. *Entonces* $\lambda = \rho(A)$, *y* $\{u\}$ *es una base para* E_λ, *donde*

$$u = \begin{pmatrix} 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}.$$

DEMOSTRACIÓN. Como $|\lambda| = \rho(A)$ las tres desigualdades de la ecuación (5) en la demostración del Teorema 5.19 son realmente igualdades, esto es

(a) $\left|\sum_{j=1}^{n} A_{kj}x_j\right| = \sum_{j=1}^{n} |A_{kj}x_j|$,

(b) $\sum_{j=1}^{n} |A_{kj}||x_j| = \sum_{j=1}^{n} |A_{kj}|\, b$,

(c) $\rho_k(A) = \rho(A)$,

donde x, b y k son los mismos que se definieron en la demostración del Teorema 5.19.

Veremos en el Ejercicio 15(b) de la Sección 7.1 que (a) se satisface si y sólo si todos los términos $A_{kj}x_j (j = 1, 2, \ldots, n)$ son múltiplos no negativos de algún número complejo no nulo z. Sin pérdida de generalidad, supondremos que $|z| = 1$. Luego existen números reales no negativos $c_1, \ldots, c_n$ tales que

$$A_{kj}x_j = c_j z. \tag{6}$$

Evidentemente (b) se satisface si y sólo si para cada j tenemos que $A_{kj} = 0$ o $|x_j| = b$. Como se supone que cada elemento de A es positivo, concluimos que (b) se satisface si y sólo si

$$|x_j| = b \qquad \text{para } j = 1, 2, \ldots, n. \tag{7}$$

Así tenemos que la ecuación (5), y por tanto el inciso (c) anterior, es válida para $k = 1, 2, \ldots, n$.

De la ecuación (6) vemos que

$$x_j = \frac{c_j}{A_{kj}} z \qquad (j = 1, 2, \ldots, n),$$

y por lo tanto, de la ecuación (7),

$$b = |x_j| = \left|\frac{c_j}{A_{kj}} z\right| = \frac{c_j}{A_{kj}} \qquad (j = 1, 2, \ldots, n).$$

Por tanto $x_j = bz$ para $j = 1, 2, \ldots, n$. Así

$$x = \begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = \begin{pmatrix} bz \\ \cdot \\ \cdot \\ \cdot \\ bz \end{pmatrix} = bz \begin{pmatrix} 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix},$$

y entonces $\{u\}$ es una base para E_λ.

Claramente se ve que u es un eigenvector de A correspondiente al eigenvalor $\rho(A)$ puesto que, de acuerdo con el inciso (c) anterior,

$$Au = A\begin{pmatrix}1\\ \cdot\\ \cdot\\ \cdot\\ 1\end{pmatrix} = \begin{pmatrix}\sum_{j=1}^{n} A_{1j}\\ \cdot\\ \cdot\\ \cdot\\ \sum_{j=1}^{n} A_{nj}\end{pmatrix} = \begin{pmatrix}\rho_1(A)\\ \cdot\\ \cdot\\ \cdot\\ \rho_n(A)\end{pmatrix} = \begin{pmatrix}\rho(A)\\ \cdot\\ \cdot\\ \cdot\\ \rho(A)\end{pmatrix} = \rho(A)\begin{pmatrix}1\\ \cdot\\ \cdot\\ \cdot\\ 1\end{pmatrix} = \rho(A)u$$

Pero el párrafo anterior muestra que si λ es cualquier valor de A tal que $|\lambda| = \rho(A)$, entonces u es un eigenvector al que corresponde λ. Por lo tanto $\lambda = \rho(A)$. ■

Corolario 1. *Sea* $A \in M_{n\times n}(C)$ *una matriz en la cual cada elemento es positivo, y sea* λ *un eigenvalor de* A *tal que* $|\lambda| = \nu(A)$. *Entonces* $\lambda = \nu(A)$, *y la dimensión de* E_λ *es* 1.

DEMOSTRACIÓN. Ejercicio.

Corolario 2. *Sea* $A \in M_{n\times n}(C)$ *una matriz de transición en la que cada elemento es positivo y sea* λ *un eigenvalor de* A *distinto de* 1. *Entonces* $|\lambda| < 1$. *Además, la dimensión del eigenespacio correspondiente al eigenvalor* 1 *es* 1.

DEMOSTRACIÓN. Ejercicio.

Nuestro siguiente resultado generaliza el corolario anterior para matrices regulares de transición y, por tanto, demuestra que las matrices regulares de transición satisfacen las dos primeras condiciones de los Teoremas 5.16 y 5.17.

Teorema 5.22. *Sea* A *una matriz regular de transición.*

(a) *Si* λ *es un eigenvalor de* A, *entonces* $|\lambda| \leq 1$.
(b) *Si* λ *es un eigenvalor de* A *tal que* $|\lambda| = 1$, *entonces* λ *es el número real* 1 *y* $dim(E_\lambda) = 1$.

En otras palabras, $\lambda = 1$ *es el único eigenvalor de* A *cuyo valor absoluto es* 1 *y* $dim(E_\lambda) = 1$. *Todos los demás eigenvalores de* A *tienen valores absolutos menores que* 1.

DEMOSTRACIÓN. El inciso (a) ya fue demostrado como Corolario 2 del Teorema 5.19.

Como A es regular, existe un entero positivo s tal que A^s tiene únicamente elementos positivos. Como A es una matriz de transición y los elementos de A^s son positivos, los elementos de $A^{s+1} = A^s(A)$ son positivos. Sea λ un eigenvalor de A cuyo valor absoluto es 1. Entonces λ^s y λ^{s+1}

son eigenvalores de A^s y A^{s+1}, respectivamente, y tienen un valor absoluto de 1, y de acuerdo con el Corolario 2 del Teorema 5.21, $\lambda^s = \lambda^{s+1} = 1$. Luego $\lambda = 1$. Sean E_λ y E'_λ los eigenespacios de A y de A^s, respectivamente, correspondientes al eigenvalor $\lambda = 1$. Entonces $E_\lambda \subseteq E'_\lambda$, pero E'_λ tiene dimensión 1 (Corolario 2 del Teorema 5.21). Por lo tanto $E_\lambda = E'_\lambda$ y $\dim(E_\lambda) = 1$. ■

Corolario. *Sea* A *una matriz regular de transición diagonalizable. Entonces existe* $\lim_{m\to\infty} A^m$.

El corolario anterior, que se obtiene directamente de los Teoremas 5.22 y 5.17, no es el mejor resultado posible. De hecho, puede demostrarse que si A es una matriz regular de transición, entonces la multiplicidad de 1 como eigenvalor de A es 1. Entonces, de acuerdo con el Teorema 5.12, se satisface la tercera condición del Teorema 5.16. Así, si A es una matriz regular de transición, existe $\lim_{m\to\infty} A^m$, sea A diagonalizable o no. Sin embargo, como ocurrió con el Teorema 5.16, el hecho de que la multiplicidad de 1 como eigenvalor de A sea 1 no puede demostrarse ahora. Sin embargo, enunciaremos este resultado aquí (dejando la demostración para el Ejercicio 20 de la Sección 6.2) y deduciremos más hechos sobre el $\lim_{m\to\infty} A^m$ cuando A es una matriz regular de transición.

Teorema 5.23. *Sea* A *una matriz regular de transición de* $n \times n$. *Entonces*

(a) *La multiplicidad de* 1 *como eigenvalor de* A *es* 1.
(b) *El* $\lim_{m\to\infty} A^m$ *existe.*
(c) $L = \lim_{m\to\infty} A^m$ *es una matriz de transición.*
(d) $AL = LA = L$.
(e) *Las columnas de* L *son idénticas. De hecho, cada columna de* L *es igual al único vector de probabilidad* v *que es también un eigenvector que corresponde al eigenvalor* 1 *de* A.
(f) *Para cualquier vector de probabilidad* x, $\lim_{m\to\infty}(A^m x) = v$.

DEMOSTRACIÓN.

(a) Véase el Ejercicio 20 de la Sección 6.2.

(b) La demostración de que $\lim_{m\to\infty} A^m$ existe se obtiene del inciso (a) y de los Teoremas 5.22 y 5.16.

(c) Como A^m es una matriz de transición de acuerdo con el corolario del Teorema 5.18, cada elemento de A^m es no negativo ($m = 1, 2, 3, \ldots$). Por tanto

$$L_{ij} = \lim_{m\to\infty} (A^m)_{ij} \geq 0 \qquad \text{para } 1 \leq i, j \leq n.$$

Además,

$$\sum_{i=1}^{n} L_{ij} = \sum_{i=1}^{n} [\lim_{m\to\infty} (A^m)_{ij}] = \lim_{m\to\infty} \left[\sum_{i=1}^{n} (A^m)_{ij} \right] = \lim_{m\to\infty} (1) = 1$$

para $1 \le j \le n$.

Luego L es una matriz de transición.

(d) De acuerdo con el Teorema 5.15,

$$AL = A(\lim_{m\to\infty} A^m) = \lim_{m\to\infty} (AA^m) = \lim_{m\to\infty} A^{m+1} = L.$$

De manera análoga $LA = L$.

(e) Como $AL = L$, de acuerdo con el inciso (d), cada una de las columnas de L es un eigenvector de A correspondiente al eigenvalor 1. Además, de acuerdo con el inciso (c), cada columna de L es un vector de probabilidad. Entonces de acuerdo con el inciso (a) cada columna de L es igual al único vector de probabilidad v correspondiente al eigenvalor 1 de A.

(f) Sea x cualquier vector de probabilidad y hágase a $y = Lx$. Entonces y es un vector de probabilidad (corolario del Teorema 5.18), y según el inciso (d) $Ay = ALx = Lx = y$. Por lo tanto y es también un eigenvector que corresponde al eigenvalor 1 de A. Entonces, de acuerdo con el inciso (e), $y = v$. ■

Definición. *El vector* v *del inciso (e) del teorema anterior se denomina* vector de probabilidad fija *(o* vector estacionario) *de la matriz regular de transición* A.

Utilizaremos ahora el Teorema 5.23 para obtener información sobre los porcentajes eventuales en cada estado de una cadena de Markov que tiene una matriz regular de transición.

Ejemplo 20. En una investigación realizada en la antigua Persia se obtuvo como resultado que en un día en particular el 50% de los persas prefería una hogaza de pan, el 30% prefería una jarra de vino y el 20% restante prefería un momento de holgorio. Una investigación subsecuente realizada un mes después arrojó los datos siguientes: De los que prefirieron una hogaza de pan en la primera investigación, el 40% siguieron prefiriendo una hogaza de pan, el 10% ahora prefirieron una jarra de vino y el 50% prefirieron irse de parranda; de los que prefirieron una jarra de vino en la primera investigación, el 20% prefirió ahora una hogaza de pan, el 70% continuaron prefiriendo una jarra de vino y el 10% ahora prefirieron la parranda; de los que prefirieron el holgorio en la primera investigación, el 20% prefirió ahora una hogaza de pan, el 20% prefirió ahora una jarra de vino y el 60% continuó prefiriendo lo mismo.

La situación descrita en el párrafo anterior es una cadena de Markov de tres estados en donde los estados son las tres posibles preferencias. Supo-

niendo que la tendencia anterior continúe, podemos predecir el porcentaje de persas en cada uno de los estados para cada mes a partir de la investigación original. Haciendo que los estados uno, dos y tres sean respectivamente las preferencias por el pan, el vino y el holgorio, vemos que el vector de probabilidad que da la probabilidad inicial de estar en cada uno de los estados es

$$P = \begin{pmatrix} 0.50 \\ 0.30 \\ 0.20 \end{pmatrix},$$

y la matriz de transición será

$$A = \begin{pmatrix} 0.40 & 0.20 & 0.20 \\ 0.10 & 0.70 & 0.20 \\ 0.50 & 0.10 & 0.60 \end{pmatrix}.$$

Las probabilidades de estar en cada uno de los estados m meses después de la encuesta original son las coordenadas del vector $A^m P$. El lector podrá verificar que

$$AP = \begin{pmatrix} 0.30 \\ 0.30 \\ 0.40 \end{pmatrix}, \qquad A^2 P = A(AP) = \begin{pmatrix} 0.26 \\ 0.32 \\ 0.42 \end{pmatrix}, \qquad A^3 P = A(A^2 P) = \begin{pmatrix} 0.252 \\ 0.334 \\ 0.414 \end{pmatrix},$$

y

$$A^4 P = A(A^3 P) = \begin{pmatrix} 0.2504 \\ 0.3418 \\ 0.4078 \end{pmatrix}.$$

Nótese la convergencia progresiva de $A^m P$.

Como A es regular, la predicción a largo plazo relativa a las preferencias de los persas puede obtenerse calculando el vector de probabilidad fija para A. Este vector es el vector de probabilidad único v tal que $(A - I)v = 0$. Si

$$v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix},$$

vemos que la ecuación matricial $(A - I)v = 0$ arroja el siguiente sistema de ecuaciones lineales:

$$\begin{cases} -0.60v_1 + 0.20v_2 + 0.20v_3 = 0 \\ 0.10v_1 - 0.30v_2 + 0.20v_3 = 0 \\ 0.50v_1 + 0.10v_2 - 0.40v_3 = 0. \end{cases}$$

Se puede demostrar fácilmente que

$$\begin{pmatrix} 5 \\ 7 \\ 8 \end{pmatrix}$$

es una base para el espacio de soluciones de este sistema. Por lo tanto el único vector de probabilidad fija para A es

$$\begin{pmatrix} \dfrac{5}{5+7+8} \\ \dfrac{7}{5+7+8} \\ \dfrac{8}{5+7+8} \end{pmatrix} = \begin{pmatrix} 0.25 \\ 0.35 \\ 0.40 \end{pmatrix}.$$

Entonces a largo plazo, el 25% de los persas preferirá una hogaza de pan, el 35% preferirá una jarra de vino y el 40% preferirá un rato de holgorio.

Nótese que si

$$Q = \begin{pmatrix} 5 & 0 & 3 \\ 7 & 1 & 1 \\ 8 & -1 & -4 \end{pmatrix},$$

entonces

$$Q^{-1}AQ = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.2 \end{pmatrix}.$$

Y así

$$\lim_{m\to\infty} A^m = Q\left[\lim_{m\to\infty}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0 \end{pmatrix}^m\right]Q^{-1} = Q\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}Q^{-1}$$

$$= \begin{pmatrix} 0.25 & 0.25 & 0.25 \\ 0.35 & 0.35 & 0.35 \\ 0.40 & 0.40 & 0.40 \end{pmatrix}.$$

Ejemplo 21. Los granjeros de Lamron siembran un cultivo por año —ya sea maíz, soya o trigo. Como ellos creen en la necesidad de hacer una rotación de cultivos, estos agricultores no siembran el mismo cultivo en años subsecuentes. De hecho, de la superficie total de cultivo en la cual se siembra un determinado cultivo, exactamente la mitad será sembrada con los otros dos cultivos durante el año siguiente. Este año, 300 hectáreas fueron sembradas de maíz, 200 hectáreas fueron sembradas de soya y 100 hectáreas de trigo.

La situación descrita en el párrafo anterior es otra cadena de Markov de tres estados en la que los tres estados corresponden respectivamente a las áreas sembradas de maíz, soya y trigo. Sin embargo, en este problema se dio la cantidad de tierra destinada a cada cultivo en vez de los porcentajes de la superficie total de cultivo (600 hectáreas). Transformando estas cantidades en fracciones de la superficie total, vemos que la matriz de transición A y el vector de probabilidad inicial de la cadena de Markov son

$$A = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \quad \text{y} \quad P = \begin{pmatrix} \frac{300}{600} \\ \frac{200}{600} \\ \frac{100}{600} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{3} \\ \frac{1}{6} \end{pmatrix}.$$

Entonces la fracción de la superficie total de cultivo destinada a cada uno de los cultivos en m años estará dada por las coordenadas de $A^m P$ y las proporciones eventuales de la superficie total de cultivo destinadas para cada cultivo son las coordenadas de $\lim_{m\to\infty} A^m P$. Así, las cantidades eventuales de tierra destinada a cada cultivo se obtienen al multiplicar este límite por la superficie total de cultivo; esto es, las cantidades de tierra para ser utilizadas eventualmente en cada cultivo son las coordenadas de $600(\lim_{m\to\infty} A^m P)$.

Como A es una matriz regular de transición, el Teorema 5.23 demuestra que $\lim_{m\to\infty} A^m$ es una matriz L en la cual cada columna es igual al único vector de probabilidad fija para A. Puede verse fácilmente que el vector de probabilidad fija para A es

$$\begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}.$$

Entonces

$$L = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix},$$

y así

$$600(\lim_{m\to\infty} A^m P) = 600LP = \begin{pmatrix} 200 \\ 200 \\ 200 \end{pmatrix}.$$

Luego, a largo plazo esperamos que 200 hectáreas sean sembradas cada año con cada cultivo. (Para un cálculo directo de $600(\lim_{m\to\infty} A^m P)$, véase el Ejercicio 14.)

En esta sección nos hemos concentrado fundamentalmente en la teoría de las matrices regulares de transición. Existe otra clase interesante de matrices de transición que puede representarse en la forma

$$\begin{pmatrix} I & B \\ O & C \end{pmatrix},$$

donde I es una matriz identidad y O es una matriz nula. (Tales matrices de transición no son regulares puesto que el bloque inferior izquierdo continúa siendo O para cualquier potencia de la matriz.) Los estados correspondientes a la submatriz identidad se denominan *estados absorbentes* porque uno de tales estados no puede ser abandonado una vez que se ha entrado en él. Una cadena de Markov se llama *cadena de Markov absorbente* si es posible pasar de un estado cualquiera a un estado absorbente en un número finito de etapas.

Obsérvese que la cadena de Markov que describió las características de los registros en un plantel de bachillerato es una cadena de Markov absorbente con los estados 1 y 4 como sus estados absorbentes. Los lectores interesados en conocer más sobre las cadenas de Markov absorbentes deben consultar *Introduction to Finite Mathematics* (tercera edición), por J. Kemeny, J. Snell y G. Thompson, Prentice-Hall, Inc., 1974 o bien *Discrete Mathematical Models* por Fred S. Roberts, Prentice-Hall, Inc., 1976.

Una aplicación

En especies que se reproducen sexualmente, las características de una progenie con respecto a una característica genética en particular están determinadas por un par de genes, uno heredado de cada padre. Los genes para una característica determinada son de dos tipos, G y g. El gen G representa la característica dominante y g representa la característica recesiva. La descendencia con genotipos GG o Gg muestran las características dominantes mientras que la progenie con características gg tienen características recesivas. Por ejemplo, en los humanos, los ojos cafés son una característica dominante y los ojos azules son la característica recesiva correspondiente; por ello, cualquier descendencia con genotipos GG o Gg tendrán los ojos cafés mientras aquellos cuyo tipo sea gg tendrán los ojos azules.

Consideremos la probabilidad de tener descendencia de cada genotipo para un progenitor macho de genotipo Gg. (Supondremos que la población bajo consideración es grande, que el cruzamiento es aleatorio con respecto al genotipo y que la distribución de cada genotipo dentro de la población es independiente del sexo y de la esperanza de vida.)

Sea

$$P = \begin{pmatrix} p \\ q \\ r \end{pmatrix}$$

la proporción de la población adulta con genotipos GG, Gg y gg, respectivamente, al principio del experimento. Este experimento describe una cadena de Markov de tres estados con una matriz de transición.

$$\begin{array}{lc} & \text{Genotipo del Progenitor} \\ & \begin{array}{ccc} \text{GG} & \text{Gg} & \text{gg} \end{array} \\ \begin{array}{l} \text{Genotipo} \\ \text{del} \\ \text{Descendiente} \end{array} \begin{array}{c} \text{GG} \\ \text{Gg} \\ \text{gg} \end{array} & \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix} = B. \end{array}$$

Se puede verificar fácilmente que B^2 contiene sólo elementos positivos; luego, B es regular. Entonces, al permitir que sólo los machos del genotipo Gg se reproduzcan, la proporción de la descendencia en la población que tiene un cierto genotipo se estabilizará en el vector de probabilidad fija para B, que es

$$\begin{pmatrix} \frac{1}{4} \\ \frac{1}{2} \\ \frac{1}{4} \end{pmatrix}.$$

Ahora supóngase que se deben realizar experimentos semejantes con machos de genotipos GG y gg. Estos experimentos son cadenas de Markov de tres estados con matrices de transición

$$A = \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{y} \quad C = \begin{pmatrix} 0 & 0 & 0 \\ 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \end{pmatrix},$$

respectivamente. Con el objeto de considerar el caso donde todos los genotipos machos pueden reproducirse, debemos formar la matriz de transición $M = pA + qB + rC$, que es una combinación lineal de A, B y C ponderada por la proporción de machos de cada genotipo. Entonces

$$M = \begin{pmatrix} p + \frac{1}{2}q & \frac{1}{2}p + \frac{1}{4}q & 0 \\ \frac{1}{2}q + r & \frac{1}{2}p + \frac{1}{2}q + \frac{1}{2}r & p + \frac{1}{2}q \\ 0 & \frac{1}{4}q + \frac{1}{2}r & \frac{1}{2}q + r \end{pmatrix}.$$

Para simplificar la notación, sea $a = p + \frac{1}{2}q$ y $b = \frac{1}{2}q + r$. (Los números a y b representan respectivamente las proporciones de los genes G y

g en la población.) Entonces

$$M = \begin{pmatrix} a & \frac{1}{2}a & 0 \\ b & \frac{1}{2} & a \\ 0 & \frac{1}{2}b & b \end{pmatrix},$$

donde $a + b = p + q + r = 1$.

Sean p', q' y r' la proporción de la descendencia de primera generación que tiene respectivamente genotipos GG, Gg y gg. Entonces

$$\begin{pmatrix} p' \\ q' \\ r' \end{pmatrix} = MP = \begin{pmatrix} ap + \frac{1}{2}aq \\ bp + \frac{1}{2}q + ar \\ \frac{1}{2}bq + br \end{pmatrix} = \begin{pmatrix} a^2 \\ 2ab \\ b^2 \end{pmatrix}.$$

Con objeto de considerar los efectos de cruzamientos no controlados entre la progenie de primera generación, se debe determinar una nueva matriz de transición $\tilde{M}$ basada en la distribución de los genotipos de primera generación. Como antes, encontramos que

$$\tilde{M} = \begin{pmatrix} p' + \frac{1}{2}q' & \frac{1}{2}p' + \frac{1}{4}q' & 0 \\ \frac{1}{2}q' + r' & \frac{1}{2}p' + \frac{1}{2}q' + \frac{1}{2}r' & p' + \frac{1}{2}q' \\ 0 & \frac{1}{4}q' + \frac{1}{2}r' & \frac{1}{2}q' + r' \end{pmatrix} = \begin{pmatrix} a' & \frac{1}{2}a' & 0 \\ b' & \frac{1}{2} & a' \\ 0 & \frac{1}{2}b' & b' \end{pmatrix},$$

donde $a' = p' + \frac{1}{2}q'$ y $b' = \frac{1}{2}q' + r'$. Pero

$$a' = a^2 + \tfrac{1}{2}(2ab) = a(a + b) = a$$

y

$$b' = \tfrac{1}{2}(2ab) + b^2 = b(a + b) = b.$$

Entonces $\tilde{M} = M$, y así la distribución de una progenie de segunda generación entre los tres genotipos es

$$\begin{aligned} \tilde{M}(MP) = M^2P &= \begin{pmatrix} a^3 + a^2b \\ a^2b + ab + ab^2 \\ ab^2 + b^3 \end{pmatrix} = \begin{pmatrix} a^2(a + b) \\ ab(a + 1 + b) \\ b^2(a + b) \end{pmatrix} = \begin{pmatrix} a^2 \\ 2ab \\ b^2 \end{pmatrix} \\ &= MP, \end{aligned}$$

que es la misma que la descendencia de primera generación. En otras palabras, MP es el vector de probabilidad fija para M, y el equilibrio genético se alcanza en la población después de sólo una generación. (Este resultado se llama *Ley de Hardy-Weinberg.)* Nótese que en el caso especial importante donde $a = b$ (o de manera equivalente, donde $p = r$), la distribución en el equilibrio es

$$MP = \begin{pmatrix} a^2 \\ 2ab \\ b^2 \end{pmatrix} = \begin{pmatrix} a^2 \\ 2a^2 \\ a^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{4} \\ \frac{1}{2} \\ \frac{1}{4} \end{pmatrix}.$$

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.
   (a) Si $A \in M_{n\times n}(C)$ y $\lim_{m\to\infty} A^m = L$, entonces, para cualquier matriz invertible $Q \in M_{n\times n}(C)$, $\lim_{m\to\infty} QA^mQ^{-1} = QLQ^{-1}$.
   (b) Si 2 es un eigenvalor de $A \in M_{n\times n}(C)$, entonces el $\lim_{m\to\infty} A^m$ no existe.
   (c) Un vector

$$\begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} \in \mathsf{R}^n$$

   tal que $x_1 + \ldots + x_n = 1$ es un vector de probabilidad.
   (d) La suma de los elementos de cada renglón de una matriz de transición es 1.
   (e) El producto de una matriz de transición por un vector de probabilidad es un vector de probabilidad.
   (f) La matriz

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ -1 & 1 & 0 \end{pmatrix}$$

   no tiene a 3 por eigenvalor.
   (g) Toda matriz de transición tiene a 1 por eigenvalor.
   (h) Ninguna matriz de transición puede tener a -1 por eigenvalor.
   (i) Si A es una matriz de transición, entonces $\lim_{m\to\infty} A^m$ existe.
   (j) Si A es una matriz regular de transición, entonces el $\lim_{m\to\infty} A^m$ existe y tiene rango 1.

2. Determinar si el $\lim_{m\to\infty} A^m$ existe o no para las siguientes matrices A. Si el límite existe, calcularlo.

(a) $\begin{pmatrix} 0.1 & 0.7 \\ 0.7 & 0.1 \end{pmatrix}$ (b) $\begin{pmatrix} 0.50 & 2 \\ 0.75 & 0 \end{pmatrix}$ (c) $\begin{pmatrix} 0.4 & 0.7 \\ 0.6 & 0.3 \end{pmatrix}$

(d) $\begin{pmatrix} 2 & 1 \\ -2 & 0 \end{pmatrix}$ (e) $\begin{pmatrix} -2 & -1 \\ 4 & 3 \end{pmatrix}$ (f) $\begin{pmatrix} -0.5 & 0.5 \\ 0.9 & -0.1 \end{pmatrix}$

(g) $\begin{pmatrix} -1.8 & 0 & -1.4 \\ -5.6 & 1 & -2.8 \\ 2.8 & 0 & 2.4 \end{pmatrix}$ (h) $\begin{pmatrix} -2.5 & 4.5 & 7.5 \\ -1.5 & -2.5 & -1.5 \\ -1.5 & 4.5 & 6.5 \end{pmatrix}$

(i) $\begin{pmatrix} -\frac{1}{2} - 2i & 4i & \frac{1}{2} + 5i \\ 1 + 2i & -3i & -1 - 4i \\ -1 - 2i & 4i & 1 + 5i \end{pmatrix}$

(j) $\begin{pmatrix} \frac{-26}{3} + \frac{i}{3} & \frac{-28}{3} - \frac{4i}{3} & 28 \\ \frac{-7}{3} + \frac{2i}{3} & \frac{-5}{3} + \frac{i}{3} & 7 - 2i \\ \frac{-13}{6} + i & \frac{-5}{6} + i & \frac{35}{6} - \frac{10i}{3} \end{pmatrix}$

3. Demostrar que si $A_1, A_2, A_3, \ldots$ es una sucesión de matrices de $n \times p$ con elementos complejos tal que $\lim_{m\to\infty} A_m = L$, entonces $\lim_{m\to\infty} A_m^t = L^t$

4. Demostrar que si $A \in M_{n\times n}(C)$ es diagonalizable y $L = \lim_{m\to\infty} A^m$ existe, entonces $L = I_n$ o bien rango$(L) < n$.

5. Encontrar las matrices A y B de 2×2 que tengan como elementos a números reales tales que $\lim_{m\to\infty} A^m$, $\lim_{m\to\infty} B^m$ y $\lim_{m\to\infty} (AB)^m$ existen pero $\lim_{m\to\infty} (AB)^m \neq (\lim_{m\to\infty} A^m)(\lim_{m\to\infty} B^m)$.

6. Una unidad de traumatología de un hospital ha determinado que al momento de llegar al hospital el 30% de sus pacientes es ambulatorio y el 70% debe guardar cama. Un mes después de su llegada, el 60% de los pacientes ambulatorios se ha recuperado, 20% permanece ambulatorio y 20% ahora debe guardar cama. Después del mismo lapso, 10% de los pacientes encamados se ha recuperado, 20% ahora es ambulatorio, 50% permanece encamado y el 20% ha muerto. Determinar el porcentaje de pacientes que se recuperaron, son ambulatorios, están encamados y han muerto un mes después de su llegada. También determinar el porcentaje eventual de pacientes de cada tipo.

7. Un jugador principia un juego de azar colocando una ficha en el casillero 2 (marcada *salida).* (Véase Fig. 5.4.) Se lanza un dado y la ficha se mueve un cuadro a la izquierda si se obtiene 1 o 2 y un cuadro a la derecha si se obtiene 3, 4, 5 o 6. Este proceso continúa hasta que la ficha llega al cuadro 1 (en cuyo caso el jugador gana el juego) o en el cuadro 4 (en cuyo caso el jugador pierde el juego). ¿Cuál es la probabilidad de ganar este juego?

Gana	Salida		Pierde
1	2	3	4

figura 5.4

8. ¿Cuáles de las matrices siguientes son matrices regulares de transición?

(a) $\begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.3 & 0.2 & 0.5 \\ 0.5 & 0.5 & 0 \end{pmatrix}$ (b) $\begin{pmatrix} 0.5 & 0 & 1 \\ 0.5 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ (c) $\begin{pmatrix} 0.5 & 0 & 0 \\ 0.5 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$

(d) $\begin{pmatrix} 0.5 & 0 & 1 \\ 0.5 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ (e) $\begin{pmatrix} \frac{1}{3} & 0 & 0 \\ \frac{1}{3} & 1 & 0 \\ \frac{1}{3} & 0 & 1 \end{pmatrix}$ (f) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.7 & 0.2 \\ 0 & 0.3 & 0.8 \end{pmatrix}$

(g) $\begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 1 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 1 \end{pmatrix}$ (h) $\begin{pmatrix} \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 1 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 1 \end{pmatrix}$

9. Calcular $\lim_{m\to\infty} A^m$, si existe, en cada una de las matrices A del Ejercicio 8.

10. Cada una de las matrices siguientes es una matriz regular de transición para una cadena de Markov de 3 estados. En todos los casos el vector de probabilidad inicial es

$$P = \begin{pmatrix} 0.3 \\ 0.3 \\ 0.4 \end{pmatrix}.$$

Para cada matriz de transición, calcular la proporción de los objetos en cada estado determinando el vector de probabilidad fija.

(a) $\begin{pmatrix} 0.6 & 0.1 & 0.1 \\ 0.1 & 0.9 & 0.2 \\ 0.3 & 0 & 0.7 \end{pmatrix}$ (b) $\begin{pmatrix} 0.8 & 0.1 & 0.2 \\ 0.1 & 0.8 & 0.2 \\ 0.1 & 0.1 & 0.6 \end{pmatrix}$ (c) $\begin{pmatrix} 0.9 & 0.1 & 0.1 \\ 0.1 & 0.6 & 0.1 \\ 0 & 0.3 & 0.8 \end{pmatrix}$

(d) $\begin{pmatrix} 0.4 & 0.2 & 0.2 \\ 0.1 & 0.7 & 0.2 \\ 0.5 & 0.1 & 0.6 \end{pmatrix}$ (e) $\begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \\ 0.3 & 0.2 & 0.5 \end{pmatrix}$ (f) $\begin{pmatrix} 0.6 & 0 & 0.4 \\ 0.2 & 0.8 & 0.2 \\ 0.2 & 0.2 & 0.4 \end{pmatrix}$

11. En 1940 una investigación municipal del uso de la tierra mostró que el 10% de la tierra del municipio era urbana, el 50% no estaba utilizada y el 40% estaba destinada a usos agrícolas. Cinco años después una investigación de actualización reveló que el 70% de la superficie urbana había permanecido urbana, 10% se había convertido en no utilizada y el 20% se había transformado en superficie agrícola. De la misma manera, 20% de la superficie no utilizada se había convertido en urbana, 60% había permanecido no utilizada y 20% se había convertido en superficie agrícola. Finalmente, la investigación de 1945 mostró que el 20% de la superficie

agrícola se había convertido en no utilizada, mientras que el 80% permaneció agrícola. Suponiendo que las tendencias indicadas por la investigación de 1945 continuaran, calcular los porcentajes de las superficies urbanas, no utilizadas y agrícolas en el municipio en 1950 y los porcentajes eventuales correspondientes.

12. Se coloca un protector de pañal en cada pañal utilizado por un bebé. Si después de un cambio de pañal el protector está sucio, entonces es desechado; de lo contrario, el protector se lava con los pañales y es reutilizado, a menos que haya sido usado ya tres veces, en cuyo caso se desecha (aun cuando nunca se haya ensuciado). La probabilidad de que un bebé ensucie un protector de pañales es de un tercio. Si al principio sólo se tienen pañales nuevos, en un tiempo cualquiera ¿qué proporción de los protectores de pañales serán nuevos, con una utilización y con dos utilizaciones?

13. En 1965 la industria automotriz determinó que el 40% de los americanos poseedores de autos conducía autos grandes, 20% conducía autos de tamaño mediano y el 40% conducía autos pequeños. Una segunda investigación en 1975 mostró que el 70% de los dueños de autos grandes de 1965 aún poseía autos grandes en 1975, pero el 30% había cambiado a autos de tamaño mediano. De aquellos que poseían automóviles de tamaño mediano en 1965, 10% había cambiado a autos grandes, el 70% seguía guiando autos medianos y el 20% había cambiado a autos pequeños en 1975. Finalmente, de los dueños de autos pequeños de 1965, el 10% poseía autos medianos y el 90% poseía autos pequeños en 1975. Suponiendo que esta tendencia continúe, determinar el porcentaje de americanos que poseerán autos de cada uno de los tamaños en 1985 y los porcentajes eventuales correspondientes.

14. Demostrar que si A es tal como en el Ejemplo 21, entonces

$$A^m = \begin{pmatrix} r_m & r_{m+1} & r_{m+1} \\ r_{m+1} & r_m & r_{m+1} \\ r_{m+1} & r_{m+1} & r_m \end{pmatrix},$$

donde

$$r_m = \frac{1}{3}\left[1 + \frac{(-1)^m}{2^{m-1}}\right].$$

Deducir que

$$600(A^m P) = A^m \begin{pmatrix} 300 \\ 200 \\ 100 \end{pmatrix} = \begin{pmatrix} 200 + \frac{(-1)^m}{2^m}(100) \\ 200 \\ 200 + \frac{(-1)^{m+1}}{2^m}(100) \end{pmatrix}.$$

15. Demostrar el Teorema 5.18 y su corolario.

16. Demostrar los dos corolarios del Teorema 5.21.

17. Demostrar el corolario del Teorema 5.22.

Definición. *Si* $A \in M_{n\times n}(C)$, *se define* $e^A = \lim_{m\to\infty} B_m$, *donde*

$$B_m = I + A + \frac{A^2}{2!} + \ldots + \frac{A^m}{m!}.$$

Luego, e^A *es la suma de la serie infinita*

$$I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \ldots,$$

y B_m *es la suma parcial* m-*ésima de esta serie. Nótese la analogía con la serie de potencias*

$$e^a = 1 + a + \frac{a^2}{2!} + \frac{a^3}{3!} + \ldots,$$

válida para todo número complejo a.

18. Calcular e^O y e^I, donde O e I son respectivamente las matrices nula e identidad de $n \times n$.

19. Supóngase que $P^{-1}AP$ es una matriz diagonal D. Demostrar que $e^A = P^{-1}e^DP$.

20. Sea $A \in M_{n\times n}(C)$ diagonalizable. Utilizar el resultado del Ejercicio 19 para demostrar que e^A existe. (El Ejercicio 21 de la Sección 6.2 demostrará que e^B existe para toda $B \in M_{n\times n}(C)$.)

21. Encontrar $A, B \in M_{2\times 2}(R)$ tal que $e^Ae^B \neq e^{A+B}$.

22. Demostrar que una función diferenciable X: $R \to R^n$ es una solución al sistema de ecuaciones diferenciales definido en el Ejercicio 17 de la Sección 5.2, si y sólo si X es de la forma $X(t) = e^{tA}v$ para alguna $v \in R^n$, donde A es como se define en dicho ejercicio.

5.4* SUBESPACIOS INVARIANTES

En la Sección 5.1 observamos que si x es un eigenvector de un operador lineal T, entonces T mapea al subespacio generado por $\{x\}$ en sí mismo.

* Esta sección no es necesaria en el Capítulo 7.

Los subespacios que se mapean en sí mismos son de una gran importancia en el estudio de los operadores lineales.

Definición. *Sea* T *un operador lineal en un espacio vectorial* V. *Un subespacio* W *de* V *se llama* subespacio T-invariante *de* V *si* $\mathsf{T}(\mathsf{W}) \subseteq \mathsf{W}$, *esto es* $\mathsf{T}(x) \in \mathsf{W}$ *para toda* $x \in \mathsf{W}$.

Para cualquier operador lineal T en V los subespacios $\{0\}$ y V son T-invariantes. Estos dos subespacios se llaman subespacios T-invariantes *impropios;* todos los demás se llaman subespacios T-invariantes *propios.*

Es deseable descomponer un espacio vectorial dimensionalmente finito V en una suma directa de tantos espacios T-invariantes propios como sea posible, puesto que el comportamiento de T puede ser inferido a través de su comportamiento en cada uno de los sumandos directos. Si T es diagonalizable, entonces V puede descomponerse en una suma directa de subespacios T-invariantes unidimensionales, que son los subespacios generados por los vectores de la base formada por los eigenvectores de T. (Véase Ejercicio 7.) En general no existe tal descomposición. En el Capítulo 6 consideraremos algunas maneras para descomponer a V en una suma directa de subespacios T-invariantes cuando T no sea diagonalizable. En esta sección estudiaremos dos propiedades básicas de las sumas directas de subespacios T-invariantes.

Para un operador lineal T en un espacio vectorial V, la restricción de T a un subespacio T-invariante W es un mapeo de W en sí mismo. (Véase Apéndice B.) Es fácil demostrar que este mapeo T_W es un operador lineal en W. (Véase Ejercicio 4.) Nuestro primer resultado relaciona el polinomio característico de T_W con el de T.

Teorema 5.24. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V, *y sea* W *un subespacio* T*-invariante de* V. *Entonces el polinomio característico de* T_W *divide al polinomio característico de* T.

DEMOSTRACIÓN. Extiéndase una base $\gamma = \{x_1, \ldots, x_k\}$ para W a una base $\beta = \{x_1, \ldots, x_k, \ldots, x_n\}$ para V. Sea $A = [\mathsf{T}]_\beta$ y $B_1 = [\mathsf{T}_\mathsf{W}]_\beta$. Entonces por el Ejercicio 5

$$A = \begin{pmatrix} B_1 & B_2 \\ O & B_3 \end{pmatrix},$$

donde O es una matriz nula de $(n-k) \times k$. Si $f(t)$ es el polinomio característico de T y $g(t)$ es el polinomio característico de T_W, entonces, de acuerdo con el Ejercicio 9 de la Sección 4.3,

$$f(t) = \det(A - tI_n) = \det\begin{pmatrix} B_1 - tI_k & B_2 \\ O & B_3 - tI_{n-k} \end{pmatrix} = g(t) \cdot \det(B_3 - tI_{n-k}).$$

Y tenemos que $g(t)$ divide a $f(t)$. ■

Ejemplo 22. Sea T: $R^4 \rightarrow R^4$ definida mediante

$$T(a, b, c, d) = (a + b + 2c - d, b + d, 2c - d, c + d),$$

y sea $W = \{(t, s, 0, 0): \ t, s \in R\}$. Obsérvese que W es un subespacio T-invariante de R^4, ya que

$$T(a, b, 0, 0) = (a + b, b, 0, 0) \in W.$$

Sea $\gamma = \{e_1, e_2\}$ y nótese que γ es una base para W. Extiéndase γ a la base estándar β para R^4. Entonces

$$B_1 = [T_W]_\gamma = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{y} \quad A = [T]_\beta = \begin{pmatrix} 1 & 1 & 2 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

de acuerdo con la notación del Teorema 5.24. Luego, si $f(t)$ es el polinomio característico de T y $g(t)$ es el polinomio característico de T_W, entonces

$$f(t) = \det(A - tI_4) = \det\begin{pmatrix} 1-t & 1 & 2 & -1 \\ 0 & 1-t & 0 & 1 \\ 0 & 0 & 2-t & -1 \\ 0 & 0 & 1 & 1-t \end{pmatrix}$$

$$= \det\begin{pmatrix} 1-t & 1 \\ 0 & 1-t \end{pmatrix} \cdot \det\begin{pmatrix} 2-t & -1 \\ 1 & 1-t \end{pmatrix} = g(t) \cdot \det\begin{pmatrix} 2-t & -1 \\ 1 & 1-t \end{pmatrix}.$$

El siguiente teorema muestra que si V es la suma directa de subespacios T-invariantes, entonces el polinomio característico de T está completamente determinado por los polinomios característicos de las restricciones de T para cada uno de los sumandos directos.

Teorema 5.25. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V, *y supóngase que* $V = W_1 \oplus \ldots \oplus W_k$, *donde* W_i *es un subespacio* T*-invariante de* V *para toda* $i (1 \leq i \leq k)$. *Si* $f(t)$ *es el polinomio característico de* T *y* $f_i(t)$ *es el polinomio característico de* $T_{W_i} (1 \leq \leq i \leq k)$, *entonces*

$$f(t) = f_1(t) \cdot f_2(t) \cdot \ldots \cdot f_k(t).$$

DEMOSTRACIÓN. La demostración se hará por inducción sobre k. Supóngase primeramente que $k = 2$. Sea β_1 una base para W_1, β_2 una base para W_2 y $\beta = \beta_1 \cup \beta_2$. Entonces β es una base para V. Sean $A = [T]_\beta$, $B_1 = [T_{W_1}]_{\beta_1}$ y $B_2 = [T_{W_2}]_{\beta_2}$. Se ve fácilmente por el Ejercicio 5 que

$$A = \begin{pmatrix} B_1 & O \\ O' & B_2 \end{pmatrix},$$

donde O y O' son matrices nulas. Por lo tanto, por el Ejercicio 9 de la Sección 4.3,

$$f(t) = \det(A - tI) = \det(B_1 - tI) \cdot \det(B_2 - tI) = f_1(t) \cdot f_2(t)$$

lo que demuestra el resultado si $k = 2$.

Ahora supóngase que el teorema es cierto para $k - 1$ sumandos, donde $k - 1$ es algún entero mayor o igual a 1. Supóngase que V es una suma directa de k sumandos,

$$\mathsf{V} = \mathsf{W}_1 \oplus \mathsf{W}_2 \oplus \ldots \oplus \mathsf{W}_k,$$

y defínase $\mathsf{W} = \mathsf{W}_1 + \mathsf{W}_2 + \ldots + \mathsf{W}_{k-1}$. Puede verificarse fácilmente que $\mathsf{V} = \mathsf{W} \oplus \mathsf{W}_k$. Así, por el caso para $k = 2$, $f(t) = g(t) \cdot f_k(t)$, donde $g(t)$ es el polinomio característico de $\mathsf{T_W}$. Claramente se tiene que $\mathsf{W} = \mathsf{W}_1 \oplus \mathsf{W}_2 \oplus \cdots \oplus \mathsf{W}_{k-1}$. Por tanto por la hipótesis de inducción tenemos que $g(t) = f_1(t) \cdot f_2(t) \cdot \ldots \cdot f_{k-1}(t)$. Entonces $f(t) = g(t) \cdot f_k(t) = f_1(t) \cdot f_2(t) \cdot \ldots \cdot f_k(t)$. ■

Si T es un operador lineal diagonalizable en un espacio vectorial n-dimensional V para el cual los eigenvalores distintos son $\lambda_1, \lambda_2, \ldots, \lambda_k$, entonces por el Teorema 5.14 $\mathsf{V} = \mathsf{E}_{\lambda_1} \oplus \mathsf{E}_{\lambda_2} \oplus \cdots \oplus \mathsf{E}_{\lambda_k}$. Se ve claramente (véase Ejercicio 3) que cada E_{λ_i} es T-invariante y la restricción de T a E_{λ_i} tiene como polinomio característico a $(\lambda_i - t)^{m_i}$, donde m_i es la multiplicidad de λ_i. Por tanto, dentro de este contexto, el teorema anterior arroja la conclusión evidente de que el polinomio característico de T es $(\lambda_i - t)^{m_1}(\lambda_2 - t)^{m_2} \ldots (\lambda_k - t)^{m_k} = (-1)^n(t - \lambda_1)^{m_1}(t - \lambda_2)^{m_1} \ldots \ldots (t - \lambda_k)^{m_k}$.

La siguiente aplicación del Teorema 5.25 sugiere otro resultado sobre las sumas directas.

Ejemplo 23. Sea $\mathsf{T}\colon \mathsf{R}^4 \to \mathsf{R}^4$ definida mediante

$$\mathsf{T}(a, b, c, d) = (2a - b, a + b, c - d, c + d),$$

y sean $\mathsf{W}_1 = \{(s, t, 0, 0)\colon s, t \in R\}$ y $\mathsf{W}_2 = \{(0, 0, s, t)\colon s, t \in R\}$. Nótese que W_1 y W_2 son ambos T-invariantes y que $\mathsf{R}^4 = \mathsf{W}_1 \oplus \mathsf{W}_2$. Sean $\beta_1 = \{e_1, e_2\}$, $\beta_2 = \{e_3, e_4\}$ y $\beta = \{e_1, e_2, e_3, e_4\}$. Entonces β_1 es una base para W_1, β_2 es una base para W_2 y β es una base para R^4. Si $B_1 = [\mathsf{T}_{\mathsf{W}_1}]_{\beta_1}$, $B_2 = [\mathsf{T}_{\mathsf{W}_2}]_{\beta_2}$, y $A = [\mathsf{T}]_\beta$, entonces

$$B_1 = \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix}, \qquad B_2 = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix},$$

y

$$A = \begin{pmatrix} B_1 & O \\ O & B_2 \end{pmatrix} = \begin{pmatrix} 2 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

Además, si $f(t)$ es el polinomio característico de T, $f_1(t)$ es el polinomio característico de T_{W_1} y $f_2(t)$ el polinomio característico de T_{W_2}, entonces

$$f(t) = \det(A - tI) = \det(B_1 - tI) \cdot \det(B_2 - tI) = f_1(t) \cdot f_2(t).$$

La matriz A del ejemplo anterior puede obtenerse uniendo las matrices B_1 y B_2 de la manera explicada en la definición siguiente.

Definición. *Sean* B_1 *y* B_2 *matrices cuadradas (no necesariamente del mismo tamaño) que tengan elementos del mismo campo. Si* B_1 *es una matriz de* $m \times m$ *y* B_2 *es una matriz de* $n \times n$, *entonces la* suma directa *de* B_1 *y* B_2, *representada por* $B_1 \oplus B_2$, *es la matriz* A *de* $(m + n) \times (m + n)$ *tal que*

$$A_{ij} = \begin{cases} (B_1)_{ij} & \textit{para } 1 \leq i, j \leq m \\ (B_2)_{(i-m),(j-m)} & \textit{para } m + 1 \leq i, j \leq n + m \\ 0 & \textit{en cualquier otro caso.} \end{cases}$$

Si $B_1, B_2, \ldots, B_k$ *son matrices cuadradas con elementos del mismo campo, entonces definimos la suma directa de* $B_1, B_2, \ldots, B_k$ *recursivamente mediante* $B_1 \oplus B_2 \oplus \ldots \oplus B_k = (B_1 \oplus B_2 \oplus \ldots \oplus B_{k-1}) \oplus B_k$. *Si* $A = B_1 \oplus B_2 \oplus \ldots \oplus B_k$, *escribiremos a menudo*

$$A = \begin{pmatrix} B_1 & O & \cdots & O \\ O & B_2 & \cdots & O \\ \vdots & \vdots & & \vdots \\ O & O & \cdots & B_k \end{pmatrix}.$$

Ejemplo 24. Sean

$$B_1 = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}, \qquad B_2 = (3) \quad \text{y} \quad B_3 = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 1 \end{pmatrix}.$$

Entonces

$$B_1 \oplus B_2 \oplus B_3 = \begin{pmatrix} 1 & 2 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

El resultado final de esta sección relaciona las sumas directas de matrices con las sumas directas de subespacios invariantes, y enuncia el caso general de la relación entre las matrices A, B_1 y B_2 del Ejemplo 23.

Teorema 5.26. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V, *y sean* $W_1, W_2, \ldots, W_k$ *subespacios* T-*invariantes de* V *tales que* $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$. *Para cada* i, *sea* β_i *una base para* W_i *y* $\beta = \beta_1 \cup \beta_2 \cup \ldots \cup \beta_k$. *Si* $A = [T]_\beta$ *y* $A_i = [T_{W_i}]_{\beta_i}$ *para* $i = 1, 2, \ldots, k$, *entonces* $A = A_1 \oplus A_2 \oplus \ldots \oplus A_k$.

DEMOSTRACIÓN. Ejercicio.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

   (a) Existen operadores lineales T que no tienen subespacios T-invariantes.
   (b) Si T es un operador lineal en un espacio vectorial dimensionalmente finito V y W es un subespacio T-invariante de V, entonces el polinomio característico de T_W divide al polinomio característico de T.
   (c) Si T es un operador lineal en un espacio vectorial dimensionalmente finito V y si V es una suma directa de subespacios T-invariantes, entonces existe una base β para V tal que $[T]_\beta$ es una suma directa de matrices.

2. Para cada uno de los siguientes operadores lineales T, determinar si el subespacio dado W es un subespacio T-invariante de V.

   (a) Sea T el operador en $V = P_3(R)$ definido mediante $T(f) = f'$, la derivada de f, y $W = P_2(R)$.
   (b) Sea T el operador en $V = P(R)$ definido mediante $T(f(x)) = xf(x)$ y $W = P_2(R)$.
   (c) Sea T el operador en $V = R^3$ definido mediante

   $$T(a_1, a_2, a_3) = (a_1 + a_2 + a_3, a_1 + a_2 + a_3, a_1 + a_2 + a_3)$$

   y

   $$W = \{(t, t, t): \ t \in R\}.$$

   (d) Sea, T el operador en el espacio vectorial V de funciones continuas de valor real en [0, 1] definido mediante

   $$T(f)(t) = \left[\int_0^1 f(x)dx\right]t \quad \text{y} \quad W = \{f \in V: \ f(t) = at + b \text{ para algunas } a, b \in R \text{ y toda } 0 \leq t \leq 1\}.$$

3. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V.

   (a) Demostrar que $\{0\}$ y V son subespacios T-invariantes de V.
   (b) Demostrar que N(T) y R(T) son subespacios T-invariantes de V.

(c) Demostrar que si λ es un eigenvalor de T, entonces E_λ es un subespacio T-invariante de V y la restricción de T a E_λ es λI.

(d) Si $W_1, W_2, \ldots, W_k$ son subespacios T-invariantes de V, demostrar que

$$\sum_{i=1}^{k} W_i \quad \text{y} \quad \bigcap_{i=1}^{k} W_i$$

son subespacios T-invariantes de V.

4. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V y sea W un subespacio T-invariante de V.

(a) Demostrar que T_W es un operador lineal en W.

(b) Demostrar que si λ es un eigenvalor de T_W entonces λ es un eigenvalor de T.

(c) Demostrar que si x es un eigenvector de T_W entonces x es un eigenvector de T.

5. Verificar que en la demostración del Teorema 5.24

$$A = \begin{pmatrix} B_1 & B_2 \\ O & B_3 \end{pmatrix}$$

y que en la demostración del Teorema 5.25

$$A = \begin{pmatrix} B_1 & O \\ O' & B_2 \end{pmatrix}.$$

6. Demostrar el Teorema 5.26.

7.* Sea T un operador lineal en un espacio vectorial dimensionalmente finito V. Demostrar que T es diagonalizable si y sólo si V es una suma directa de subespacios T-invariantes unidimensionales.

8. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V y sean $W_1, W_2, \ldots, W_k$ subespacios T-invariantes propios de V tales que $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$. Demostrar que $\det(T) = \det(T_{W_1}) \cdot \det(T_{W_2}) \cdot \cdots \cdot \det(T_{W_k})$.

9. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V, y supóngase que $W_1, W_2, \ldots, W_k$ son subespacios T-invariantes de V tales que $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$. Demostrar que si T_{W_i} es diagonalizable para cada i, entonces T es diagonalizable.

10. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V.

(a) Demostrar que si el polinomio característico de T se descompone como un producto de factores de grado 1, lo mismo ocurre con el polinomio

característico de la restricción de T a cualquier subespacio T-invariante de V.

(b) Deducir que si el polinomio característico de T se descompone en un producto de factores de grado 1, entonces cualquier subespacio T-invariante no nulo de V contiene un eigenvector de T.

(c) Sea W cualquier subespacio T-invariante de V. Demostrar que si λ es un eigenvalor de T_W entonces el eigenespacio de T_W correspondiente a λ es $E_\lambda \cap W$, donde E_λ es el eigenespacio de T correspondiente a λ.

(d) Sea W cualquier subespacio T-invariante de V. Demostrar que si T es diagonalizable, también lo es T_W. *Sugerencia:* Utilizar los incisos (a) y (c) y el Teorema 5.14.

11. (a) Demostrar un recíproco del Ejercicio 19(a) de la Sección 5.2: Si T y U son operadores lineales diagonalizables en un espacio vectorial dimensionalmente finito tales que UT = TU, entonces T y U son simultáneamente diagonalizables. *Sugerencia:* Sean $\lambda_1, \lambda_2, \ldots, \lambda_k$ los distintos eigenvalores de T y sean E_{λ_i} $(i = 1, 2, \ldots, k)$ los correspondientes eigenespacios de T. Demuéstrese que cada E_{λ_i} es U-invariante y utilícese el Ejercicio 10(d) para obtener una base para E_{λ_i} formada por eigenvectores de U.

(b) Enunciar y demostrar la versión matricial del inciso (a).

12. (a) El resultado del Ejercicio 11(a) puede ser generalizado de la manera siguiente: Sea V un espacio vectorial dimensionalmente finito. Se dice que una colección $\mathcal{C}$ de operadores lineales diagonalizables en V es *simultáneamente diagonalizable* si existe una β para V tal que $[T]_\beta$ es una matriz diagonal para cada $T \in \mathcal{C}$.

Demostrar que una colección $\mathcal{C}$ de operadores lineales diagonalizables en un espacio vectorial dimensionalmente finito V es simultáneamente diagonalizable si y sólo si UT = TU para toda $T, U \in \mathcal{C}$. *Sugerencia:* En el caso de que UT = TU para toda $T, U \in \mathcal{C}$, establézcase primero el resultado cuando cada operador en $\mathcal{C}$ tenga únicamente un eigenvalor. Luego establézcase el resultado general mediante inducción sobre dim(V) utilizando el hecho de que V puede expresarse como la suma directa de los eigenespacios para algún operador en $\mathcal{C}$.

(b) Enunciar y demostrar la versión matricial del inciso (a).

Los Ejercicios 13 y 14 requieren que el lector esté familiarizado con el Ejercicio 29 de la Sección 1.3.

13.* Sea T un operador lineal en V y sea W un subespacio T-invariante de V. Defínase

$$\bar{T}: V/W \longrightarrow V/W \quad \text{por} \quad \bar{T}(v + W) = T(v) + W.$$

(a) Demostrar que $\bar{T}$ está bien definido; esto es, demostrar que $\bar{T}(v + W)$ es independiente de la selección de v en el coconjunto de $v + W$.

(b) Demostrar que $\bar{T}$ es un operador lineal en V/W.

(c) Defínase

$$\eta\colon\ V \to V/W \quad \text{por} \quad \eta(v) = v + W.$$

Demostrar que η es una transformación lineal con espacio nulo W y rango V/W.

(d) Demostrar que el diagrama de la Fig. 5.5 es conmutativo; esto es $\eta T = \bar{T}\eta$.

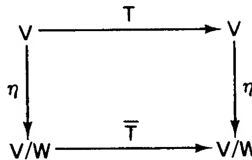


figura 5.5

14. (a) Sea T un operador lineal en un espacio vectorial dimensionalmente finito V, y sea W un subespacio T-invariante propio de V. Sean $f(t)$, $g(t)$ y $h(t)$ los polinomios característicos de T, T_W y $\bar{T}$ (tal como se definió en el Ejercicio 13), respectivamente. Demostrar que $f(t) = g(t)h(t)$. *Sugerencia:* Extender una base $\gamma = \{x_1, \ldots, x_k\}$ para W a una base $\beta = \{x_1, \ldots, x_k, \ldots, x_n\}$ para V. Mostrar que $[T]_\beta$ es de la forma

$$\begin{pmatrix} B_1 & B_2 \\ O & B_3 \end{pmatrix},$$

que $\bar{\beta} = \{x_{k+1} + W, \ldots, x_n + W\}$, es una base para V/W, y que $[\bar{T}]_{\bar{\beta}} = B_3$.

(b) Utilizar el Ejercicio 13 para demostrar que si T es diagonalizable, también lo será $\bar{T}$.

*5.5** *EL TEOREMA DE CAYLEY HAMILTON*

En la Sección 5.4 mencionamos que si T es un operador lineal en un espacio vectorial dimensionalmente finito V, entonces es deseable descomponer a V en una suma directa de tantos subespacios T-invariantes como sea posible. Cuando el polinomio característico de T se descompone en un producto de factores de grado 1, demostraremos en la Sección 6.1 que V puede

* El material de esta sección es necesario para las Secciones 5.6 y 6.3, pero no para las Secciones 6.1 y 6.2.

descomponerse siempre en una suma directa de los "eigenespacios generalizados" de T. De hecho, cuando T es diagonalizable, ésta es precisamente la descomposición dada en el Teorema 5.14. Sin embargo, si el polinomio característico de T no se descompone en un producto de factores de grado 1, entonces T puede incluso no tener eigenvalores. En este caso no podemos esperar descomponer a V en una suma directa como antes. No obstante, una descomposición de V en subespacios T-invariantes es aún posible. En esta sección definiremos el tipo especial de subespacios T-invariantes necesarios para esta descomposición y los emplearemos para demostrar uno de los más famosos teoremas del álgebra lineal, el teorema de Cayley-Hamilton.

Definiciones. *Sea* T *un operador lineal en un espacio vectorial* V. *Un subespacio* W *de* V *se llama* subespacio T-cíclico *si existe un elemento* $x \in W$ *tal que* W *es igual al subespacio generado por* $\{x, T(x), T^2(x), \ldots\}$. *En este caso decimos que* W *es* generado *por* x. *El subespacio* T-*cíclico generado por* x *será representado por* C_x.

Ejemplo 25. Sea T: $R^3 \to R^3$ definido por

$$T(a, b, c) = (-b + c, a + c, 3c).$$

Determinaremos el subespacio T-cíclico generado por $e_1 = (1, 0, 0)$. Como $T(e_1) = T(1, 0, 0) = (0, 1, 0) = e_2$ y $T^2(e_1) = T(T(e_1)) = T(e_2) = (-1, 0, 0) = -e_1$, entonces $C_{e_1} = L(\{e_1, T(e_1), T^2(e_1), \ldots\}) = L(\{e_1, e_2\}) = \{(s, t, 0)\colon\ s, t \in R\}$. Nótese que para esta transformación $C_{e_1} = C_{e_2}$.

Ejemplo 26. Sea T el operador lineal en $P(R)$ definido mediante $T(f) = f'$. Entonces C_{x^2} es igual al subespacio generado por $\{x^2, 2x, 2\}$ e igual a $P_2(R)$.

Puede verse fácilmente que los subespacios T-cíclicos son T-invariantes. Nuestro siguiente resultado establece algunas propiedades adicionales de los subespacios T-cíclicos.

Teorema 5.27. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V, *y sea* W *el subespacio* T-*cíclico de* V *generado por* $x \in V$. *Supóngase que* $dim(W) = k \geq 1$ *(y por tanto* $x \neq 0$*). Entonces*

(a) $\{x, T(x), T^2(x), \ldots, T^{k-1}(x)\}$ *es una base para* W.

(b) *Si* $-a_0, -a_1, \ldots, -a_{k-1}$ *son los escalares dados por (a) tales que* $T^k(x) = -a_0x - a_1T(x) - \ldots - a_{k-1}T^{k-1}(x)$ *entonces* $f(t) = (-1)^k(a_0 + a_1t + \ldots + a_{k-1}t^{k-1} + t^k)$ *es el polinomio característico de* T_W.

DEMOSTRACIÓN. Sea j el entero más pequeño para el que $\{x, \mathrm{T}(x), \ldots, \mathrm{T}^j(x)\}$ es linealmente independiente. (Tal j debe existir porque W es dimensionalmente finito.) Como $x \neq 0$, $j \geq 1$. Por tanto $\{x, \mathrm{T}(x), \ldots, \mathrm{T}^{j-1}(x)\}$ es linealmente independiente y $\mathrm{T}^j(x) \in L(\{x, \mathrm{T}(x), \ldots \mathrm{T}^{j-1}(x)\})$ de acuerdo con el lema del Teorema 1.10. Demostraremos por inducción matemática que $\mathrm{T}^s(x)$ está dentro de este subespacio para cualquier entero no negativo s. Esto es evidente para $0 \leq s \leq j$. Supóngase que $\mathrm{T}^m(x)$ pertenece al subespacio generado por $\{(x, \mathrm{T}(x), \ldots, \mathrm{T}^{j-1}(x)\}$ para alguna $m \geq j$. Entonces existen escalares $b_0, b_1, \ldots, b_{j-1}$ tales que

$$\mathrm{T}^m(x) = b_0 x + b_1 \mathrm{T}(x) + \ldots + b_{j-1}\mathrm{T}^{j-1}(x).$$

Aplicando T a ambos lados de la igualdad anterior, obtenemos

$$\mathrm{T}^{m+1}(x) = b_0 \mathrm{T}(x) + b_1 \mathrm{T}^2(x) + \ldots + b_{j-1}\mathrm{T}^{j}(x).$$

Luego $\mathrm{T}^{m+1}(x)$ es una combinación lineal de $\mathrm{T}(x), \mathrm{T}^2(x), \ldots, \mathrm{T}^j(x)$, cada uno de los cuales pertenece al subespacio generado por $\{x, \mathrm{T}(x), \ldots, \mathrm{T}^{j-1}(x)\}$. Entonces $\mathrm{T}^{m+1}(x)$ pertenece a este subespacio completando la inducción. Por tanto

$$\mathrm{W} = L(\{x, \mathrm{T}(x), \mathrm{T}^2(x), \ldots,\}) \subseteq L(\{x, \mathrm{T}(x), \ldots, \mathrm{T}^{j-1}(x)\}).$$

Pero claramente se ve que la inclusión recíproca es también cierta y por lo tanto $\{x, \mathrm{T}(x), \ldots, \mathrm{T}^{j-1}(x)\}$ genera a W. Como este conjunto es también linealmente independiente, es una base para W. Pero $\dim(\mathrm{W}) = k$; de manera que este conjunto debe contener k elementos. Por lo tanto $j = k$ y entonces $\{x, \mathrm{T}(x), \ldots, \mathrm{T}^{k-1}(x)\}$ es una base para W, lo que demuestra el inciso (a).

Para demostrar (b), sea $\beta = \{x, \mathrm{T}(x), \ldots, \mathrm{T}^{k-1}(x)\}$ la base del inciso (a) y sean $-a_0, -a_1, \ldots, -a_{k-1}$ escalares tales que $\mathrm{T}^k(x) = -a_0 x - a_1\mathrm{T}(x) - \ldots - a_{k-1}\mathrm{T}^{k-1}(x)$. Obsérvese que

$$[\mathrm{T_W}]_\beta = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & -a_{k-2} \\ 0 & 0 & \cdots & 1 & -a_{k-1} \end{pmatrix},$$

y entonces el polinomio característico de $[\mathrm{T_W}]_\beta$ es

$$f(t) = (-1)^k(a_0 + a_1 t + \ldots + a_{k-1}t^{k-1} + t^k)$$

por el Ejercicio 10, por lo que $f(t)$ es el polinomio característico de $\mathrm{T_W}$ demostrando así el inciso (b). ■

Definición. *En la demostración del Teorema 5.27, la matriz*

$$[\mathsf{T_W}]_\beta = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & -a_{k-2} \\ 0 & 0 & \cdots & 1 & -a_{k-1} \end{pmatrix},$$

se llama matriz compañera *del polinomio*

$$f(t) = (-1)^k(a_0 + a_1t + \ldots + a_{k-1}t^{k-1} + t^k).$$

Ahora podemos demostrar el célebre teorema de Cayley-Hamilton. El lector deberá consultar el Apéndice E para la definición de $f(\mathsf{T})$ cuando T es un operador lineal y $f(t)$ un polinomio.

Teorema 5.28. *(Cayley-Hamilton.) Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V *y sea* $f(t)$ *el polinomio característico de* T. *Entonces* $f(\mathsf{T}) = \mathsf{T}_0$ *(la transformación nula); o sea que* T *satisface a su polinomio característico.*

DEMOSTRACIÓN. Debemos demostrar que $f(\mathsf{T})(x) = 0$ para toda $x \in \mathsf{V}$. Si $x = 0$ entonces $f(\mathsf{T})(x) = 0$ puesto que $f(\mathsf{T})$ es una transformación lineal. Supóngase entonces que $x \neq 0$ y sea $\mathsf{W} = \mathsf{C}_x$. Si $\dim(\mathsf{W}) = k$, entonces, de acuerdo con el Teorema 5.27, existen escalares $-a_0, -a_1, \ldots, -a_{k-1}$ tales que

$$\mathsf{T}^k(x) = -a_0x - a_1\mathsf{T}(x) - \ldots - a_{k-1}\mathsf{T}^{k-1}(x).$$

Por lo tanto el Teorema 5.27 implica que

$$g(t) = (-1)^k(a_0 + a_1t + \ldots + a_{k-1}t^{k-1} + t^k)$$

es el polinomio característico de $\mathsf{T_W}$. Combinando estas dos ecuaciones tenemos

$$g(\mathsf{T})(x) = (-1)^k(a_0\mathsf{I} + a_1\mathsf{T} + \ldots + a_{k-1}\mathsf{T}^{k-1} + \mathsf{T}^k)(x) = 0.$$

De acuerdo con el Teorema 5.24, $g(t)$ divide a $f(t)$; por lo tanto existe un polinomio $q(t)$ tal que $f(t) = q(t)g(t)$, y entonces

$$f(\mathsf{T})(x) = q(\mathsf{T})g(\mathsf{T})(x) = q(\mathsf{T})(g(\mathsf{T})(x)) = q(\mathsf{T})(0) = 0. \quad \blacksquare$$

Ejemplo 27. Sea $\mathsf{T}: \mathsf{R}^2 \to \mathsf{R}^2$ definida mediante $\mathsf{T}(a, b) = (a + 2b, -2a + b)$ y sea $\beta = \{e_1, e_2\}$. Entonces

$$A = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix},$$

donde $A = [T]_\beta$. Por lo tanto el polinomio característico de T es

$$f(t) = \det(A - tI) = \det\begin{pmatrix} 1-t & 2 \\ -2 & 1-t \end{pmatrix} = t^2 - 2t + 5.$$

Se puede verificar fácilmente que $T_0 = f(T) = T^2 - 2T + 5$. Del mismo modo

$$f(A) = A^2 - 2A + 5I = \begin{pmatrix} -3 & 4 \\ -4 & -3 \end{pmatrix} + \begin{pmatrix} -2 & -4 \\ 4 & -2 \end{pmatrix} + \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Ejemplo 28. Sea T el operador lineal en $P_2(R)$ definido mediante $T(f) = f' + f$. Se ve fácilmente que el polinomio característico de T es $g(t) = (1 - t)^3 = -t^3 + 3t^2 - 3t + 1$. Ahora bien,

$$T^2(f) = T(f' + f) = f'' + 2f' + f$$

y

$$T^3(f) = f''' + 3f'' + 3f' + f.$$

Por lo tanto

$$g(T)(f) = -T^3(f) + 3T^2(f) - 3T(f) + I(f) = -f'''.$$

Pero para $f \in P_2(R)$, $f''' = 0$. Por lo tanto $g(T) = T_0$.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

(a) Sean C_x y C_y subespacios T-cíclicos de un operador lineal T en un espacio vectorial dimensionalmente finito V. Si $C_x = C_y$, entonces $x = y$.

(b) Si T es un operador lineal en un espacio vectorial dimensionalmente finito, entonces $C_x = C_{T(x)}$.

(c) Sea T un operador lineal en un espacio vectorial n-dimensional. Existe un polinomio $g(t)$ de grado n tal que $g(T) = T_0$.

(d) El polinomio característico de la matriz compañera de $g(t) = (-1)^k(a_0 + a_1t + \ldots + a_{k-1}t^{k-1} + t^k)$ es $g(t)$.

(e) Un polinomio de la forma $(-1)^k(a_0 + a_1t + \ldots + a_{k-1}t^{k-1} + t^k)$ es el polinomio característico de algún operador lineal.

2. Encontrar una base para el subespacio T-cíclico C_z en cada uno de los incisos siguientes.

(a) $T\colon R^4 \to R^4$ definida mediante $T(a_1, a_2, a_3, a_4) = (a_1 + a_2, a_2 - a_3, a_1 + a_3, a_1 + a_4)$ y $z = e_1$.

(b) $T: P_3(R) \to P_3(R)$ definida mediante $T(f) = f''$ y $z = x^3$.
(c) $T: M_{2\times 2}(R) \to M_{2\times 2}(R)$ definida mediante

$$T(A) = A^t \quad \text{y} \quad z = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

(d) $T: M_{2\times 2}(R) \to M_{2\times 2}(R)$ definida mediante

$$T(A) = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} A \quad \text{y} \quad z = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

3. Para cada uno de los operadores lineales T del Ejercicio 2.

(i) Calcular el polinomio característico de T_{C_z}.
(ii) Calcular el polinomio característico de T.
(iii) Verificar el teorema de Cayley-Hamilton para T.

4. Sea $T: V \to V$ un operador lineal, demostrar que para cualquier $x \in V$, C_x es T-invariante.

5. Demostrar el teorema de Cayley-Hamilton para matrices: Si A es una matriz de $n \times n$, de polinomio característico $f(t)$, entonces $f(A) = 0$ es la matriz nula de $n \times n$.

6. Sea V un espacio vectorial bidimensional y sea $T: V \to V$ un operador lineal. Demostrar que V es un subespacio T-cíclico de sí mismo o bien $T = \lambda I$ para algún escalar λ.

7. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V, y sea $f(t) = (-1)^n t^n + a_{n-1}t^{n-1} + \ldots + a_1 t + a_0$ el polinomio característico de T. Demostrar que

(a) T es invertible si y sólo si $a_0 \neq 0$.
(b) Si T es invertible, entonces

$$T^{-1} = \frac{(-1)^{n+1}}{a_0} T^{n-1} - \frac{a_{n-1}}{a_0} T^{n-2} - \ldots - \frac{a_1}{a_0} I.$$

8. Sea T un operador lineal en un espacio vectorial V, y sea W un subespacio T-invariante de V. Demostrar que para cualquier polinomio $g(t)$, W es $g(T)$-invariante.

9. Sea T un operador lineal en un espacio vectorial V. Demostrar para cualquier $x \in V$, que el subespacio T-cíclico C_x es el subespacio T-invariante más pequeño de V que contiene a x; esto es, para cualquier subespacio T-invariante W que contenga a x, $C_x \subseteq W$.

10. Sea A la matriz de $k \times k$

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & \cdots & 0 & -a_{k-2} \\ 0 & 0 & \cdots & 1 & -a_{k-1} \end{pmatrix},$$

donde $a_0, a_1, \ldots, a_{k-1}$ son escalares cualesquiera. Demostrar que el polinomio característico de A es

$$(-1)^k(a_0 + a_1 t + \ldots + a_{k-1}t^{k-1} + t^k).$$

Sugerencia: Utilizar inducción sobre k.

11. Utilizar el Ejercicio 22 de la Sección 5.1 para obtener una demostración fácil del teorema de Cayley-Hamilton para operadores diagonalizables.

12. Sea T un operador lineal en un espacio vectorial n-dimensional V. Demostrar que el subespacio generado por $\{I, T, T^2, \ldots\}$ es un subespacio de $\mathcal{L}(V)$ cuya dimensión no excede a n.

5.6* EL POLINOMIO MINIMO

Para un operador dado T en un espacio vectorial dimensionalmente finito V, el teorema de Cayley-Hamilton muestra que existe un polinomio $f(t)$ para el cual $f(T) = T_0$, que es el polinomio característico de T. Existen muchos otros polinomios que tienen esta propiedad. Uno de los más importantes, el polinomio mínimo, proporciona otro medio para estudiar a los operadores lineales.

Definición. *Sea* T *un operador lineal en un espacio vectorial* V. *Un polinomio* p(t) *se llama* polinomio mínimo *para* T *si* p(t) *es un polinomio mónico de grado positivo mínimo para el cual* $p(T) = T_0$. *(Recuérdese del Apéndice E que un polinomio mónico es aquel en el cual el coeficiente principal es* 1.)

Es fácil ver que cualquier operador lineal T en un espacio vectorial dimensionalmente finito tiene un polinomio mínimo. Nótese que si $g(t)$ es un polinomio de grado k con coeficiente principal a tal que $g(T) = T_0$, entonces $h(t) = (1/a)g(t)$ es un polinomio mónico de grado k para el que $h(T) = T_0$. Por lo tanto el teorema de Cayley-Hamilton muestra que

* Esta sección se requiere únicamente para la Sección 6.3.

el grado de un polinomio mínimo para T es a lo más igual a la dimensión del espacio vectorial en el que T está definido. El resultado siguiente muestra que el requerimiento de que un polinomio mínimo sea mónico garantiza que sea único.

Teorema 5.29. *Sea* p(t) *un polinomio mínimo para un operador lineal* T *en un espacio vectorial dimensionalmente finito* V.

(a) *Si* g(t) *es un polinomio cualquiera para el cual* $g(T) = T_0$, *entonces* p(t) *divide a* g(t). *En particular,* p(t) *divide al polinomio característico de* T.

(b) *Existe únicamente un polinomio mínimo para* T; *esto es,* p(t) *es único.*

DEMOSTRACIÓN.

(a) Sea $g(t)$ un polinomio cualquiera para el cual $g(T) = T_0$. El algoritmo de la división para polinomios (véase Apéndice E) implica que existen polinomios $q(t)$ y $r(t)$ tales que

$$g(t) = q(t)p(t) + r(t),$$

donde $r(t)$ es de grado menor que $p(t)$. Sustituyendo a T en la ecuación (8) y usando el hecho de que $g(T) = p(T) = T_0$, tenemos que $r(T) = T_0$. Como $r(t)$ es de grado menor que $p(t)$ y $p(t)$ es un polinomio mínimo, $r(t)$ debe ser el polinomio nulo. Luego la ecuación (8) se convierte en $g(t) = q(t)p(t)$ demostrando (a).

(b) Supóngase que $p_1(t)$ y $p_2(t)$ son cada uno polinomios mínimos para T. Entonces de acuerdo con el inciso (a) $p_1(t)$ divide a $p_2(t)$. Pero como $p_1(t)$ y $p_2(t)$ tienen el mismo grado no negativo, debe de tenerse que $p_1(t) = kp_2(t)$ para algún escalar no nulo k. Además, como $p_1(t)$ y $p_2(t)$ son mónicos, $k = 1$, por lo que $p_1(t) = p_2(t)$. ■

Antes de continuar con nuestro estudio del polinomio mínimo para un operador, introduciremos el concepto de polinomio mínimo para una matriz.

Definición. *El* polinomio mínimo p(t) *para* $A \in M_{n \times n}(F)$ *es el polinomio mónico de grado positivo mínimo para el que* p(A) *es igual a la matriz nula.*

A lo largo de este libro, los enunciados sobre transformaciones lineales han sido traducidos en enunciados sobre matrices y viceversa. El siguiente teorema y su corolario son de este tipo.

Teorema 5.30. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V *y sea* β *una base para* V. *Entonces el polinomio mínimo para* T *es el mismo que el polinomio mínimo para* $[T]_\beta$.

DEMOSTRACIÓN. Ejercicio.

Corolario. *Para cualquier* $A \in M_{n\times n}(F)$, *el polinomio mínimo para* A *es el mismo que el polinomio mínimo para* L_A.

DEMOSTRACIÓN. Ejercicio.

Como consecuencia del teorema anterior y su corolario, los teoremas siguientes de esta sección que se enuncien para operadores son también válidos para matrices.

En el resto de esta sección estudiaremos principalmente polinomios mínimos para operadores cuyos polinomios característicos se descomponen en un producto de factores de grado 1. En la Sección 6.3 se hará un estudio más detallado de los polinomios mínimos.

Teorema 5.31. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V, *y sea* p(t) *el polinomio mínimo para* T. *Un escalar* λ *es un eigenvalor de* T *si y sólo si* $p(\lambda) = 0$. *Por lo tanto el polinomio característico y el polinomio mínimo para* T *tienen los mismos ceros.*

DEMOSTRACIÓN. Sea $f(t)$ el polinomio característico de T. Como $p(t)$ divide a $f(t)$, $f(t) = q(t)p(t)$ para algún polinomio $q(t)$. Sea λ un cero de $p(t)$. Entonces

$$f(\lambda) = q(\lambda)p(\lambda) = q(\lambda) \cdot 0 = 0.$$

Por lo que λ es también un cero para $f(t)$; esto es, λ es un eigenvalor de T.

Recíprocamente, supóngase que λ es un eigenvalor de T y sea $x \in V$ un eigenvector correspondiente a λ. Entonces, por el Ejercicio 22 de la Sección 5.1, tenemos que

$$0 = T_0(x) = p(T)(x) = p(\lambda)x.$$

Como $x \neq 0$, $p(\lambda) = 0$ y por lo tanto λ es un cero para $p(t)$. ■

Como una consecuencia inmediata del resultado anterior tenemos el corolario siguiente.

Corolario. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V *con polinomio mínimo* p(t) *y con polinomio característico* f(t). *Supóngase que* f(t) *se factoriza como*

$$f(t) = (\lambda_1 - t)^{n_1}(\lambda_2 - t)^{n_2} \ldots (\lambda_k - t)^{n_k},$$

donde $\lambda_1, \lambda_2, \ldots, \lambda_k$ *son los distintos eigenvalores de* T. *Entonces existen enteros* $m_1, m_2, \ldots, m_k$ *tales que* $1 \leq m_i \leq n_i$ *para toda* i *y*

$$p(t) = (t - \lambda_1)^{m_1}(t - \lambda_2)^{m_2} \ldots (t - \lambda_k)^{m_k}.$$

Ejemplo 29. Calcularemos el polinomio mínimo para la matriz

$$A = \begin{pmatrix} 3 & -1 & 0 \\ 0 & 2 & 0 \\ 1 & -1 & 2 \end{pmatrix}.$$

Como el polinomio característico de A es

$$f(t) = \det\begin{pmatrix} 3-t & -1 & 0 \\ 0 & 2-t & 0 \\ 1 & -1 & 2-t \end{pmatrix} = -(t-2)^2(t-3),$$

el polinomio mínimo para A debe ser, de acuerdo con el corolario del Teorema 5.31 $(t-2)(t-3)$, o bien $(t-2)^2(t-3)$. Al sustituir A en $p(t) = (t-2)(t-3)$ se demuestra que $p(A)$ es la matriz nula; luego, $p(t)$ es el polinomio mínimo de A.

Ejemplo 30. Sea T: $\mathsf{R}^2 \to \mathsf{R}^2$ definida mediante

$$\mathsf{T}(a, b) = (2a + 5b, 6a + b).$$

Si β es la base estándar para R^2, entonces

$$[\mathsf{T}]_\beta = \begin{pmatrix} 2 & 5 \\ 6 & 1 \end{pmatrix}.$$

Así el polinomio característico de $[\mathsf{T}]_\beta$, y por tanto de T, es

$$f(t) = \det\begin{pmatrix} 2-t & 5 \\ 6 & 1-t \end{pmatrix} = (t-7)(t+4).$$

Y entonces el polinomio mínimo para T debe ser $(t-7)(t+4)$.

Ejemplo 31. Sea D: $\mathsf{P}_2(R) \to \mathsf{P}_2(R)$ el operador de diferenciación definido mediante $\mathsf{D}(f) = f'$. Calcularemos el polinomio mínimo para D. Para la base $\beta = \{1, t, t^2\}$ tenemos que

$$[\mathsf{D}]_\beta = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Por lo tanto el polinomio característico de D es $-t^3$, y el corolario al Teorema 5.31 muestra que el polinomio mínimo para D es t, t^2 o t^3. Como $\mathsf{D}^2(t^2) = 2 \neq 0$, $\mathsf{D}^2 \neq \mathsf{T}_0$, por lo que el polinomio mínimo para D debe ser t^3.

En el ejemplo anterior es fácil verificar que $\mathsf{P}_2(R)$ es un subespacio D-cíclico (de sí mismo). En este ejemplo vimos que los polinomios mínimo y característico fueron del mismo grado y esto no es ninguna coincidencia.

Teorema 5.32. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V. *Si* V *es un subespacio* T*-cíclico, o sea, si* $V = C_x$ *para alguna* $x \in V$, *entonces el polinomio característico* $f(t)$ *y el polinomio mínimo* $p(t)$ *para* T *son del mismo grado. Por lo tanto* $f(t) = (-1)^n p(t)$.

DEMOSTRACIÓN. Si V es un subespacio T-cíclico, entonces existe un elemento $x \in V$ tal que

$$\beta = \{x, T(x), \ldots, T^{n-1}(x)\}$$

es una base para V (Teorema 5.27). Sea

$$g(t) = a_0 + a_1 t + \ldots + a_k t^k,$$

donde $a_k \neq 0$ y $0 \leq k \leq n$. Entonces

$$g(T)(x) = a_0 x + a_1 T(x) + \ldots + a_k T^k(x)$$

es una combinación lineal de elementos de β que tienen al menos un coeficiente no nulo, el cual supondremos que es a_k. Como β es linealmente independiente, $g(T)(x) \neq 0$ y por lo tanto $g(T) \neq T_0$. Y así el polinomio mínimo para T es de grado n, que es también el grado del polinomio característico de T. ■

El Teorema 5.32 enuncia una condición bajo la cual el grado del polinomio mínimo para un operador es tan grande como posible. Investigaremos ahora cuándo el grado del polinomio mínimo es tan pequeño como posible. Se tiene del Teorema 5.31 que si el polinomio característico de un operador con k eigenvalores distintos se descompone como el producto de factores de grado 1, entonces el polinomio mínimo debe ser al menos de grado k. El siguiente teorema muestra que los operadores para los cuales el grado del polinomio mínimo es tan pequeño como posible son precisamente los operadores diagonalizables.

Teorema 5.33. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V. *Entonces* T *es diagonalizable si y sólo si el polinomio mínimo para* T *es de la forma*

$$p(t) = (t - \lambda_1)(t - \lambda_2) \ldots (t - \lambda_k),$$

donde $\lambda_1, \lambda_2, \ldots, \lambda_k$ *son escalares distintos. (Nótese que* $\lambda_1, \lambda_2, \ldots, \lambda_k$ *son necesariamente los distintos eigenvalores de* T.)

DEMOSTRACIÓN. Supóngase que T es diagonalizable y sean $\lambda_1, \lambda_2, \ldots, \lambda_k$ los distintos eigenvalores de T con eigenespacios correspondientes $E_{\lambda_1}, E_{\lambda_2}, \ldots, E_{\lambda_k}$. Si $\dim(E_{\lambda_i}) = n_i (i = 1, 2, \ldots, k)$, entonces el polinomio característico de T, $f(t)$, es

$$f(t) = (\lambda_1 - t)^{n_1}(\lambda_2 - t)^{n_2} \ldots (\lambda_k - t)^{n_k}.$$

Sea $p(t)$ el polinomio mínimo de T y defínase

$$q(t) = (t - \lambda_1)(t - \lambda_2) \ldots (t - \lambda_k).$$

Demostraremos que $q(\mathsf{T}) = \mathsf{T}_0$. Como de acuerdo con el Teorema 5.31 $q(t)$ divide a $p(t)$, se tendrá que $p(t) = q(t)$. Recuérdese que para cualquier $i(1 \leq i \leq k) x \in \mathsf{E}_{\lambda_i}$ si y sólo si $(\mathsf{T} - \lambda_i \mathsf{I})(x) = 0$. Por lo tanto para cualquier $x \in \mathsf{E}_{\lambda_i}$

$$q(\mathsf{T})(x) = (\mathsf{T} - \lambda_1\mathsf{I})(\mathsf{T} - \lambda_2\mathsf{I}) \ldots (\mathsf{T} - \lambda_k\mathsf{I})(x) = 0. \qquad (9)$$

Como T es diagonalizable, $\mathsf{V} = \mathsf{E}_{\lambda_1} \oplus \mathsf{E}_{\lambda_2} \oplus \cdots \oplus \mathsf{E}_{\lambda_k}$ por el Teorema 5.14. Así, de la ecuación (9), $q(\mathsf{T})(x) = 0$ para toda $x \in \mathsf{V}$ y por lo tanto $q(\mathsf{T}) = \mathsf{T}_0$.

Recíprocamente, supóngase que existen escalares $\lambda_1, \lambda_2, \ldots, \lambda_k$ (necesariamente eigenvalores de T) tales que el polinomio mínimo, $p(t)$, para T se factoriza en

$$p(t) = (t - \lambda_1)(t - \lambda_2) \ldots (t - \lambda_k).$$

Si $k = 1$, entonces $(\mathsf{T} - \lambda_1\mathsf{I})(x) = 0$ para toda $x \in \mathsf{V}$. Por lo tanto $\mathsf{T} = \lambda_1\mathsf{I}$, que es claramente diagonalizable. Supóngase entonces que $k > 1$. Sean $\{p_j(t): j = 1, 2, \ldots, k\}$ los polinomios de Lagrange asociados con $\lambda_1, \lambda_2, \ldots, \lambda_k$ (tal como se definieron en la Sección 1.6). La fórmula de interpolación de Lagrange muestra que

$$\sum_{j=1}^{k} p_j(t) = 1,$$

donde el lado izquierdo de la igualdad es el polinomio constante 1. Por lo tanto

$$\sum_{j=1}^{k} p_j(\mathsf{T})(x) = \mathsf{I}(x) = x \qquad (10)$$

para toda $x \in \mathsf{V}$. Además, la definición de polinomio de Lagrange muestra que $(t - \lambda_j)p_j(t) = cp(t)$ donde c es un escalar. Por lo tanto

$$(\mathsf{T} - \lambda_j\mathsf{I})p_j(\mathsf{T})(x) = cp(\mathsf{T})(x) = c\mathsf{T}_0(x) = 0,$$

de modo que $p_j(\mathsf{T})(x) \in \mathsf{E}_{\lambda_j}$. Entonces por la ecuación (10)

$$\mathsf{V} = \mathsf{E}_{\lambda_1} + \mathsf{E}_{\lambda_2} + \ldots + \mathsf{E}_{\lambda_k},$$

y así V está generado por su conjunto de eigenvectores. Por lo tanto V tiene una base de eigenvectores (Ejercicio 11 de la Sección 1.6), y entonces T, de acuerdo con el Teorema 5.4, es diagonalizable. ■

Ejemplo 32. Determinaremos todas las matrices $A \in \mathsf{M}_{2\times 2}(R)$ para las cuales $A^2 - 3A + 2I = O$ donde O es la matriz nula de 2×2. Defínase $g(t) = t^2 - 3t + 2 = (t - 1)(t - 2)$. Como $g(A) = O$, el polinomio mínimo $p(t)$ para A divide a $g(t)$. Por lo tanto los únicos candidatos

posibles para $p(t)$ son $t-1$, $t-2$ o $(t-1)(t-2)$. Nótese que en cualquiera de estos casos A es diagonalizable en virtud del Teorema 5.33. Si $p(t) = t-1$ o $p(t) = t-2$, entonces $A = I$ o $A = 2I$. Si $p(t) = (t-1)(t-2)$, entonces A es similar a

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Ejemplo 33. Demostraremos que si A es una matriz real de $n \times n$ tal que $A^3 = A$, entonces A es diagonalizable. Nótese que si $g(t) = t^3 - t = t(t+1)(t-1)$, entonces $g(A) = O$ donde O es la matriz nula de $n \times n$. Por lo tanto el polinomio mínimo $p(t)$ para A divide a $g(t)$. Como $g(t)$ no tiene factores repetidos, tampoco los tiene $p(t)$. Luego, de acuerdo con el Teorema 5.33, A es diagonalizable.

Ejemplo 34. En el Ejemplo 31 vimos que el polinomio mínimo para el operador de diferenciación D: $\mathsf{P}_2(R) \to \mathsf{P}_2(R)$ es t^3. Por tanto D no es diagonizable (Teorema 5.33).

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas. Suponer en lo que sigue que todos los espacios vectoriales son dimensionalmente finitos.

(a) Todo operador lineal T tiene un polinomio $p(t)$ de grado máximo para el cual $p(\mathsf{T}) = \mathsf{T}_0$.

(b) Todo operador lineal tiene un polinomio mínimo único.

(c) El polinomio característico de un operador lineal divide al polinomio mínimo para ese operador.

(d) Los polinomios mínimo y característico de cualquier operador diagonalizable son idénticos.

(e) Sean T un operador lineal en un espacio vectorial n-dimensional V, $p(t)$ el polinomio mínimo para T, y $f(t)$ el polinomio característico de T. Si $f(t)$ se descompone en un producto de factores de grado 1, entonces $f(t)$ divide a $[p(t)]^n$.

(f) El polinomio mínimo para un operador lineal siempre tiene el mismo grado que el polinomio característico del operador.

(g) Un operador lineal es diagonalizable si su polinomio mínimo se descompone en un producto de factores de grado 1.

(h) Sea T un operador lineal en V. Si V es un subespacio T-cíclico, entonces el grado del polinomio mínimo para T es igual a $\dim(\mathsf{V})$.

2. Calcular el polinomio mínimo para las siguientes matrices.

(a) $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ (b) $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ (c) $\begin{pmatrix} 4 & -14 & 5 \\ 1 & -4 & 2 \\ 1 & -6 & 4 \end{pmatrix}$.

3. Calcular el polinomio mínimo para cada uno de los siguientes operadores lineales.

(a) $T: P_2(R) \to P_2(R)$, donde $T(f) = f' + 2f$
(b) $T: R^2 \to R^2$, donde $T(a, b) = (a + b, a - b)$
(c) $T: M_{n\times n}(R) \to M_{n\times n}(R)$, donde $T(A) = A'$. *Sugerencia:* Nótese que $T^2 = I$.

4. Determinar cuáles de las matrices y operadores de los Ejercicios 2 y 3 son diagonalizables.

5. Describir todos los operadores lineales T en R^2 tales que T sea diagonalizable y $T^3 - 2T^2 + T = T_0$.

6. Demostrar el Teorema 5.30 y su corolario.

7. Demostrar el corolario del Teorema 5.31.

8. Sea T un operador lineal en un espacio vectorial dimensionalmente finito. Demostrar que si $g(t)$ es el polinomio mínimo de T, entonces

(a) T es invertible si y sólo si $g(0) \neq 0$.
(b) Si T es invertible y $g(t) = t^n + a_{n-1}t^{n-1} + \ldots + a_1t + a_0$, entonces

$$T^{-1} = -\left(\frac{1}{a_0}T^{n-1} + \frac{a_{n-1}}{a_0}T^{n-2} + \ldots + \frac{a_1}{a_0}I\right).$$

9. Sea T un operador lineal diagonalizable en un espacio vectorial dimensionalmente finito V. Demostrar que V es un subespacio T-cíclico si y sólo si cada uno de los eigenespacios de T es unidimensional.

10. Sea $g(t)$ el polinomio auxiliar de una ecuación diferencial lineal homogénea con coeficientes constantes (tal como se definió en la Sección 2.7), y sea V el espacio de soluciones de la ecuación diferencial. Demostrar que

(a) V es un subespacio D-invariante, donde $D: C^\infty \to C^\infty$ es el operador de diferenciación.
(b) El polinomio mínimo para D_V (la restricción de D a V) es $g(t)$.
(c) Si el grado de $g(t)$ es n, entonces el polinomio característico de $D: V \to V$ es $(-1)^n g(t)$.
Sugerencia: Para (b) y (c), utilizar el Teorema 2.36.

11. Sea $D: P(R) \to P(R)$ el operador de diferenciación en el espacio de todos los polinomios sobre R. Demostrar que no existe ningún polinomio $g(t)$ para el que $g(D) = T_0$. Por lo tanto $D: P(R) \to P(R)$ no tiene polinomio mínimo.

12. Sea V un espacio vectorial dimensionalmente finito y sea T un operador lineal en V. Supóngase que W_1 y W_2 son subespacios T-invariantes de V tales que $V = W_1 \oplus W_2$ y sean $p_1(t)$ y $p_2(t)$ los polinomios mínimos para T_{W_1} y T_{W_2}, las restricciones de T a W_1 y W_2, respectivamente. Demostrar positiva o negativamente que $p_1(t)p_2(t)$ es el polinomio mínimo para T.

13.* Sea T un operador lineal en un espacio vectorial dimensionalmente finito V y sea W_1 un subespacio T-invariante de V. Si $x \in V$ y $x \notin W_1$, demostrar lo siguiente:

(a) Existe un polinomio mónico único $g_1(t)$ de grado positivo mínimo tal que $g_1(T)(x) \in W_1$.

(b) Si $h(t)$ es un polinomio para el cual $h(T)(x) \in W_1$, entonces $g_1(t)$ divide a $h(t)$.

(c) Sea W_2 un subespacio T-invariante de V tal que $W_2 \subseteq W_1$. Demostrar que si $g_2(t)$ es el único polinomio mónico de grado positivo mínimo tal que $g_2(T)(x) \in W_2$, entonces $g_1(t)$ divide a $g_2(t)$. Deducir que $g_1(t)$ divide a los polinomios mínimo y característico de T.

Definición. *Sea* $T: V \to V$ *un operador lineal en un espacio vectorial dimensionalmente finito* V. *Para cada* x *no nula en* V *el* T-*aniquilador de* x *es el polinomio mónico* $p_x(t)$ *de menor grado positivo para el que* $p_x(T)(x) = 0$. *Obsérvese que por el Ejercicio 13(a) anterior (con* $W_1 = \{0\}$*) cualquier* x *no nula en* V *tiene un* T-*aniquilador único.*

14.* Sea T un operador lineal en un espacio vectorial dimensionalmente finito V y sea x un elemento de V no nulo.

(a) Mostrar que si $q(t)$ es un polinomio cualquiera tal que $g(T)(x) = 0$, entonces $p_x(t)$, el T-aniquilador de x, divide a $q(t)$.

(b) Sea $W = C_x$ el subespacio T-cíclico de V generado por x. Demostrar que el polinomio mínimo para T_W es $p_x(t)$ y por lo tanto la dimensión de C_x es igual, por el Teorema 5.32, al grado del T-aniquilador de x.

(c) Demostrar que $p_x(t)$ es de grado 1 si y sólo si x es un eigenvector de T.

INDICE DE DEFINICIONES PARA EL CAPITULO 5

Capítulo 6

Formas canónicas

Vimos en el Capítulo 5 que para un operador lineal T en un espacio vectorial dimensionalmente finito V, es benéfico descomponer a V en una suma directa de tantos subespacios T-invariantes propios como sea posible. El Ejercicio 7 de la Sección 5.4 muestra que V puede descomponerse en una suma directa de subespacios T-invariantes unidimensionales si y sólo si T es diagonalizable. En este capítulo estudiaremos especialmente la descomposición de V en subespacios T-invariantes propios cuando T no sea diagonalizable. Las Secciones 6.1 y 6.2 están destinadas al estudio de los operadores cuyos polinomios característicos se descomponen en un producto de factores de grado 1 y la Sección 6.3 está dedicada a los operadores lineales cuyos polinomios característicos no pueden factorizarse de esta manera. Estas descomposiciones conducirán a representaciones sencillas (canónicas) de tales operadores.

6.1 EIGENVECTORES GENERALIZADOS

En las primeras dos secciones de este capítulo consideraremos operadores lineales en espacios vectoriales dimensionalmente finitos para los que el polinomio característico se descompone en un producto de factores de grado 1. (En particular, si V es un espacio vectorial dimensionalmente finito sobre un campo algebraicamente cerrado, todo operador lineal en V satisface esta condición.) Tales operadores tienen al menos un eigenvalor. Si $\lambda_1, \lambda_2, \ldots, \lambda_n$ son los eigenvalores de T: $V \to V$ (no necesariamente diferentes), recuérdese del Teorema 5.4 que T es diagonalizable si y sólo si existe una base ordenada para V que esté formada por eigenvectores de T. Si $\beta = \{x_1, x_2, \ldots, x_n\}$ es dicha base en la cual x_j es un eigenvector correspondiente al eigenvalor λ_j, entonces

$$[T]_\beta = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Aun cuando no todo operador lineal T en V es diagonalizable, demostraremos que para cualquier operador lineal cuyo polinomio característico se descomponga en un producto de factores de grado 1 existe una base ordenada β para V tal que

$$[\mathsf{T}]_\beta = J_1 \oplus J_2 \oplus \cdots \oplus J_k = \begin{pmatrix} J_1 & O & \cdots & O \\ O & J_2 & \cdots & O \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ O & O & \cdots & J_k \end{pmatrix}.$$

donde J_i es una matriz cuadrada de la forma (λ_j) o bien de la forma

$$\begin{pmatrix} \lambda_j & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_j & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & \lambda_j & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_j \end{pmatrix}$$

para algún eigenvalor λ_j de T. A tal matriz J_i se le llamará un *bloque de Jordan* correspondiente a λ_j y la matriz $[\mathsf{T}]_\beta = J_1 \oplus J_2 \oplus \cdots \oplus J_k$ será denominada *forma canónica de Jordan* de T. Diremos también que la base ordenada β es una *base canónica de Jordan* para T. Obsérvese que cada bloque de Jordan J_i es "casi" una matriz diagonal —de hecho, $[\mathsf{T}]_\beta$ es una matriz diagonal si y sólo si cada J_i es de la forma (λ_j).

Por ejemplo, la matriz de 8×8

$$J = J_1 \oplus J_2 \oplus J_3 \oplus J_4 = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

es una forma canónica de Jordan de un operador lineal $\mathsf{T}\colon \mathsf{C}^8 \to \mathsf{C}^8$; esto es, existe una base $\beta = \{x_1, x_2, \ldots, x_8\}$ para C^8 tal que $[\mathsf{T}]_\beta = J$. Nótese que el polinomio característico de T y J es $\det(J - tI) = (t-2)^4(t-3)^2t^2$ y así la multiplicidad de cada eigenvalor es el número de veces que el eigenvalor aparece en la diagonal de J. Obsérvese también que de los vectores $x_1, x_2, \ldots, x_8$ únicamente x_1, x_4, x_5 y x_7 (los vectores de la

base correspondientes a la primera columna de cada uno de los bloques de Jordan J_1, J_2, J_3 y J_4) son eigenvectores de T.

Aunque se demostrará que todo operador cuyo polinomio característico se descompone en un producto de factores de grado 1 tiene una forma canónica de Jordan única (de acuerdo con el orden de los bloques de Jordan), no significa que la forma canónica de Jordan queda completamente determinada por el polinomio característico de la transformación. Por ejemplo, el polinomio característico de

$$J' = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

es también $(t-2)^4(t-3)^2t^2$.

Considérese de nuevo a la matriz anterior J. Hemos visto que x_1 y x_4 son eigenvectores correspondientes al eigenvalor $\lambda_1 = 2$, pero ni x_2 ni x_3 son eigenvectores. Por lo tanto $(\mathsf{T} - 2\mathsf{I})(x_1) = (\mathsf{T} - 2\mathsf{I})(x_4) = 0$, mientras que $(\mathsf{T} - 2\mathsf{I})(x_2) \neq 0$ y $(\mathsf{T} - 2\mathsf{I})(x_3) \neq 0$. Pero como $[\mathsf{T}]_\beta = J$, $\mathsf{T}(x_2) = x_1 + 2x_2$ y $\mathsf{T}(x_3) = x_2 + 2x_3$. Entonces

$$(\mathsf{T} - 2\mathsf{I})^2(x_2) = (\mathsf{T} - 2\mathsf{I})(\mathsf{T}(x_2) - 2x_2) = (\mathsf{T} - 2\mathsf{I})(x_1) = 0,$$

y, de manera semejante,

$$(\mathsf{T} - 2\mathsf{I})^3(x_3) = (\mathsf{T} - 2\mathsf{I})^2(\mathsf{T}(x_3) - 2x_3) = (\mathsf{T} - 2\mathsf{I})^2(x_2) = 0.$$

Luego, aun cuando $(\mathsf{T} - 2\mathsf{I})(x_2) \neq 0$ y $(\mathsf{T} - 2\mathsf{I})(x_3) \neq 0$, $(\mathsf{T} - 2\mathsf{I})^p(x_2) = (\mathsf{T} - 2\mathsf{I})^p(x_3) = 0$ si $p \geq 2$. Esta observación da lugar a la siguiente definición.

Definición. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V. *Un elemento no nulo* $\mathrm{x} \in \mathsf{V}$ *se llama* eigenvector generalizado *de* T *si existe un escalar* λ *tal que* $(\mathsf{T} - \lambda\mathsf{I})^{\mathrm{p}}(\mathrm{x}) = 0$ *para algún entero positivo* p. *Diremos que* x *es un eigenvector generalizado correspondiente a* λ.

Obsérvese que si x es un eigenvector generalizado de T correspondiente a λ, entonces λ es un eigenvalor de T. Ahora si p es el menor entero positivo tal que $(\mathsf{T} - \lambda\mathsf{I})^p(x) = 0$, entonces $y = (\mathsf{T} - \lambda\mathsf{I})^{p-1}(x)$ es un eigenvector de T correspondiente al eigenvalor λ.

Se ve fácilmente que si β es una base canónica de Jordan para un operador T en un espacio vectorial dimensionalmente finito V, entonces β está formada de eigenvectores generalizados de T. El Teorema 6.4 mos-

trará que una base canónica de Jordan existe para todo operador V cuyo polinomio característico se descompone en un producto de factores de grado 1. La demostración de este teorema requerirá de terminología adicional que introduciremos a continuación.

Definiciones. *Sea* T *un operador lineal en un espacio vectorial* V *y sea* x *un eigenvector generalizado de* T *correspondiente al eigenvalor* λ. *Si* p *es el entero positivo más pequeño tal que* $(\mathsf{T}-\lambda\mathsf{I})^{p}(x) = 0$, *entonces el conjunto ordenado*

$$\{(\mathsf{T}-\lambda\mathsf{I})^{p-1}(x), (\mathsf{T}-\lambda\mathsf{I})^{p-2}(x), \ldots, (\mathsf{T}-\lambda\mathsf{I})(x), x\}$$

se llama un ciclo *de eigenvectores generalizados de* T *que corresponden a* λ. *Los elementos* $(\mathsf{T}-\lambda\mathsf{I})^{p-1}(x)$ *y* x *se llaman* vector inicial *y* vector terminal *del ciclo, respectivamente. Diremos también que la* longitud *del ciclo es* p.

Recordando la matriz J de la p. 322 vemos que $\beta_1 = \{x_1, x_2, x_3\}$, $\beta_2 = \{x_4\}$, $\beta_3 = \{x_5, x_6\}$ y $\beta_4 = \{x_7, x_8\}$ son ciclos de eigenvectores generalizados de T correspondientes respectivamente a los eigenvalores 2, 2, 3 y 0. Sea W_i el subespacio generado por β_i para $1 \leq i \leq 4$. Como $\mathsf{T}(x_1) = 2x_1$, $\mathsf{T}(x_2) = x_1 + 2x_2$ y $\mathsf{T}(x_3) = x_2 + 2x_3$, W_1 es un subespacio T-invariante. De la misma manera W_2, W_3 y W_4 son subespacios T-invariantes. Se ve fácilmente que $[\mathsf{T}_{\mathsf{W}_i}]_{\beta_i} = J_i$ $(1 \leq i \leq 4)$.

Nuestro primer resultado contiene varios resultados útiles sobre ciclos.

Teorema 6.1. *Sea* T *un operador lineal en* V *y sea* γ *un ciclo de eigenvectores generalizados de* T *que corresponden al eigenvalor* λ.

(a) *El vector inicial de* γ *es un eigenvector de* T *correspondiente al eigenvalor* λ *y ningún otro miembro de* γ *es un eigenvector de* T.
(b) γ *es linealmente independiente.*
(c) *Sea* β *una base ordenada para* V. *Entonces* β *es una base canónica de Jordan para* V *si y sólo si* β *es una unión disjunta de ciclos de eigenvectores generalizados de* T.

DEMOSTRACIÓN. Demostraremos únicamente a (b); las demostraciones para (a) y (c) se dejarán como ejercicios. La demostración se hará mediante inducción sobre la longitud del ciclo γ. Si γ tiene longitud 1, entonces $\gamma = \{x_1\}$ es linealmente independiente puesto que x_1, que es un eigenvector generalizado, es un vector no nulo. Ahora supóngase que los ciclos de longitud $k-1$ son linealmente independientes para algún entero $k-1 \geq 1$. Supóngase que $\gamma = \{x_1, x_2, \ldots, x_k\}$ es un ciclo de eigenvectores generalizados que corresponden al eigenvalor λ y que

$$\sum_{i=1}^{k} a_i x_i = 0$$

para algunos escalares $a_1, a_2, \ldots, a_k$. Aplicando $T - \lambda I$ a la ecuación anterior nos da

$$\sum_{i=2}^{k} a_i x_{i-1} = 0.$$

Pero la suma en la igualdad anterior es una combinación lineal de elementos de un ciclo $\{x_1, x_2, \ldots, x_{k-1}\}$ de longitud $k - 1$. Por lo tanto $a_i = 0$ para $i = 2, 3, \ldots, k$, por lo que

$$\sum_{i=1}^{k} a_i x_i = 0$$

se reduce a $a_1 x_1 = 0$. Pero como $x_1 \neq 0$ se tiene que $a_1 = 0$, de manera que $a_1 = a_2 = \ldots = a_k = 0$, demostrando que γ es linealmente independiente. Esto completa la inducción. ■

Recuérdese que si T es un operador lineal diagonalizable en V, entonces V es la suma directa de los eigenespacios de T (Teorema 5.14). Uno de los principales resultados de esta sección (Teorema 6.5) demostrará que si T es un operador lineal cualquiera en V cuyo polinomio característico se descompone en un producto de factores de grado 1, entonces V es la suma directa de los "eigenespacios generalizados" de T (los que se definen en seguida). Por lo tanto, como los eigenespacios de un operador diagonalizable proporcionaron una base para V formada por eigenvectores, los eigenespacios generalizados de un operador darán una base canónica de Jordan para V.

Definición. *Sea λ un eigenvalor de un operador lineal* T *en* V. *El* eigenespacio generalizado *de* T *correspondiente a λ y denotado por K_λ es el conjunto*

$$K_\lambda = \{x \in V\colon\ (T - \lambda I_V)^p(x) = 0 \text{ para algún entero positivo } p\}.$$

Luego, K_λ consta del vector nulo y de todos los eigenvectores generalizados correspondientes a λ.

Nuestro siguiente teorema contiene dos hechos simples sobre los eigenespacios generalizados.

Teorema 6.2. *Sea λ un eigenvalor de un operador lineal* T *en* V. *Entonces K_λ es un subespacio de* V *que contiene a E_λ (el eigenespacio de* T *correspondiente a λ).*

DEMOSTRACIÓN. Es evidente que $0 \in K_\lambda$. Supóngase que $x, y \in K_\lambda$; entonces existen enteros positivos p y q tales que $(T - \lambda I)^p(x) = 0$ y $(T - \lambda I)^q(y) = 0$. Ahora bien

$$\begin{aligned}(T - \lambda I)^{p+q}(x + y) &= (T - \lambda I)^{p+q}(x) + (T - \lambda I)^{p+q}(y) \\ &= (T - \lambda I)^q(T - \lambda I)^p(x) + (T - \lambda I)^p(T - \lambda I)^q(y) \\ &= (T - \lambda I)^q(0) + (T - \lambda I)^p(0) \\ &= 0 + 0 = 0,\end{aligned}$$

y entonces $x + y \in K_\lambda$. Finalmente, para cualquier escalar c,

$$(T - \lambda I)^p(cx) = c(T - \lambda I)^p(x) = c0 = 0,$$

tal que $cx \in K_\lambda$. Por lo tanto K_λ es un subespacio de V.
Es evidente que $E_\lambda = N(T - \lambda I) \subseteq K_\lambda$. ■

Veremos en el Teorema 6.5 que K_λ es de hecho un subespacio T-invariante y que V es la suma directa de los eigenespacios generalizados de T. El teorema siguiente demuestra parte de un resultado posterior.

Teorema 6.3. *Sean $\lambda_1, \lambda_2, \ldots, \lambda_k$ los distintos eigenvalores de un operador lineal T en V. Entonces*

$$K_{\lambda_i} \cap \left(\sum_{j \neq i} K_{\lambda_j}\right) = \{0\} \qquad \textit{para } i = 1, 2, \ldots, k.$$

DEMOSTRACIÓN. Por conveniencia de notación consideraremos sin pérdida de generalidad que $i = 1$. Supóngase que

$$a_2x_2 + \ldots + a_kx_k = x_1 \tag{1}$$

donde $x_j \in K_{\lambda_j}$ para $1 \leq j \leq k$. Sea $p_j (1 \leq j \leq k)$ el entero positivo más pequeño tal que $(T - \lambda_j I)^{p_j}(x_j) = 0$. Supóngase que $x_1 \neq 0$; entonces $(T - \lambda_1 I)^{p_1-1}(x_1)$ es un eigenvector correspondiente al eigenvalor λ_1. Aplicando $(T - \lambda_1 I)^{p_1-1}(T - \lambda_2 I)^{p_2} \ldots (T - \lambda_k I)^{p_k}$ a ambos lados de la ecuación (1) se tiene

$$\begin{aligned} 0 &= (T - \lambda_1 I)^{p_1-1}(T - \lambda_2 I)^{p_2} \ldots (T - \lambda_k I)^{p_k}(x_1) \\ &= (T - \lambda_2 I)^{p_2} \ldots (T - \lambda_k I)^{p_k}((T - \lambda_1 I)^{p_1-1}(x_1)) \\ &= (\lambda_1 - \lambda_2)^{p_2} \ldots (\lambda_1 - \lambda_k)^{p_k}(T - \lambda_1 I)^{p_1-1}(x_1) \end{aligned}$$

en virtud del Ejercicio 22 de la Sección 5.1. Por lo tanto dado que $\lambda_1, \lambda_2, \ldots, \lambda_k$ son distintos, $(T - \lambda_1 I)^{p_1-1}(x_1) = 0$, contradiciendo el que $(T - \lambda_1 I)^{p_1-1}(x_1)$ sea un eigenvector. Concluimos que $x_1 = 0$. ■

Corolario. *Ningún vector puede ser un eigenvector generalizado correspondiente a eigenvalores diferentes del mismo operador.*

Estamos ahora preparados para demostrar la existencia de una forma canónica de Jordan para cada operador lineal en un espacio vectorial dimensionalmente-finito cuyo polinomio característico se descomponga en un producto de factores de grado 1.

Lema. *Sea T un operador lineal en un espacio vectorial dimensionalmente finito V. Sea $S_i (1 \leq i \leq k)$ un ciclo de eigenvectores generalizados de T correspondientes al eigenvalor λ y sean p_i y y_i la longitud y el vector inicial de*

S_i, *respectivamente. Si* $\{y_1, y_2, \ldots, y_k\}$ *es un conjunto linealmente independiente que contiene* k *elementos, entonces*

$$\bigcup_{i=1}^{k} S_i$$

es un conjunto linealmente independiente que contiene

$$\sum_{i=1}^{k} p_i$$

elementos.

DEMOSTRACIÓN. Supóngase sin pérdida de generalidad que $p_1 \geq p_2 \geq \ldots \ldots \geq p_k$. La demostración se hará por inducción sobre p_1. Si $p_1 = 1$, entonces $p_1 = \ldots = p_k = 1$. Por lo tanto, cada ciclo S_i contiene un solo elemento, de modo que

$$\bigcup_{i=1}^{k} S_i = \{y_1, y_2, \ldots, y_k\}$$

es un conjunto linealmente independiente que por hipótesis contiene

$$\sum_{i=1}^{k} p_i = k$$

elementos.

Ahora supóngase que el teorema es cierto siempre que $p_1 < n$ y sea $S_i (1 \leq i \leq k)$ un ciclo de eigenvectores generalizados de T correspondientes al eigenvalor λ con longitud p_i y vector inicial y_i. Supóngase que $n = p_1 \geq p_2 \geq \ldots \geq p_k$ y sea $r(1 < r \leq k)$ el subíndice más grande tal que $p_r > 1$. Sea

$$S = \bigcup_{i=1}^{k} S_i,$$

y sea $S'_i (1 \leq i \leq r)$ el ciclo obtenido al suprimir el vector terminal x_i de S_i. Entonces S'_i es un ciclo de eigenvectores generalizados que corresponden al eigenvalor λ con longitud $p_i - 1$ y vector inicial y_i. Como $\{y_1, \ldots, y_r\}$ es linealmente independiente, se tiene por la hipótesis de inducción que

$$S' = \bigcup_{i=1}^{r} S'_i$$

es un conjunto linealmente independiente que contiene

$$\sum_{i=1}^{r} (p_i - 1)$$

elementos. Es evidente que

$$\bigcup_{i=1}^{k} S_i = S' \cup \{x_1, \ldots, x_k\}$$

es una unión disjunta. Entonces

$$\bigcup_{i=1}^{k} S_i$$

contiene

$$\sum_{i=1}^{r} (p_i - 1) + k = \sum_{i=1}^{k} (p_i - 1) + k = \sum_{i=1}^{k} p_i$$

elementos.

Necesitamos demostrar únicamente que

$$S = \bigcup_{i=1}^{k} S_i$$

es un conjunto linealmente independiente. Supóngase que para algunos escalares a_z

$$\sum_{z \in S} a_z z = 0. \tag{2}$$

Como, de acuerdo con el Teorema 6.1(a), y_i es un eigenvector de T correspondiente al eigenvalor λ, $(\mathsf{T} - \lambda\mathsf{I})(y_i) = 0$ para $1 \le i \le k$. Por lo tanto, al aplicar $\mathsf{T} - \lambda\mathsf{I}$ a ambos lados de la ecuación (2), obtenemos

$$0 = \sum_{z \in S} a_z(\mathsf{T} - \lambda\mathsf{I})(z) = \sum_{z \in Z} a_z(\mathsf{T} - \lambda\mathsf{I})(z), \tag{3}$$

donde $Z = \{v \in S \colon v \neq y_i \text{ para } 1 \le i \le k\}$. Pero la suma final en la ecuación (3) es una combinación lineal de elementos de S'; luego como S' es linealmente independiente, se tiene que $a_z = 0$ si $z \in Z$. Luego la ecuación (2) se reduce a una combinación lineal de $\{y_1, y_2, \ldots, y_k\}$, la cual es, por hipótesis, linealmente independiente. Por lo tanto, todos los coeficientes a_z de la ecuación (2) son iguales a cero, demostrando que S es linealmente independiente. ■

Ejemplo 1. Sea $\mathsf{T}\colon \mathsf{C}^{11} \to \mathsf{C}^{11}$ un operador lineal cuyo polinomio característico se descompone en un producto de factores de grado 1. Este ejemplo y el Ejercicio 8 ilustran la manera cómo una base canónica de Jordan γ para la restricción de T a $\mathsf{R}(\mathsf{T})$ se extiende a una base canónica de Jordan β para T en el Caso 1 del Teorema 6.4. Supóngase que $\gamma = \{y_1, w_1, y_2, v_1, v_2, v_3\}$ y que

$$[\mathsf{T}_1]_\gamma = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix},$$

donde T_1 denota la restricción de T a R(T). Entonces $\{y_1, w_1\}$, $\{y_2\}$, $\{v_1, v_2\}$ y $\{v_3\}$, que son los ciclos que componen a γ, corresponden respectivamente a los eigenvalores 0, 0, 2 y 3. En la notación del Teorema 6.4, γ_0, la unión de los ciclos correspondientes a 0, es igual a $\{y_1, w_1\} \cup \{y_2\}$. Entonces $Y = \{y_1, y_2\}$ (el conjunto que contiene a los eigenvectores contenidos en γ que corresponden a cero) es un subconjunto linealmente independiente de N(T) que puede ser extendido a una base $Y \cup Z = \{y_1, y_2, z_1, z_2, z_3\}$ para N(T). Finalmente, escójase a x_1 y x_2 tales que $T(x_1) = w_1$ y $T(x_2) = y_2$. Entonces $\{y_1, w_1, x_1\}$, $\{y_2, x_2\}$, $\{z_1\}$, $\{z_2\}$ y $\{z_3\}$ son ciclos de eigenvectores generalizados de T correspondientes al eigenvalor cero y $\beta = \{y_1, w_1, x_1, y_2, x_2, v_1, v_2, v_3, z_1, z_2, z_3\}$ es una base ordenada para C^{11}. Obsérvese que

$$[T]_\beta = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Teorema 6.4. *Sea* T *un operador lineal en un espacio vectorial* n*-dimensional* V *tal que el polinomio característico de* T *se descompone en un producto de factores de grado* 1. *Entonces existe una base canónica de Jordan para* T; *esto es, existe una base ordenada* β *para* V *que es una unión disjunta de ciclos de eigenvectores generalizados de* T.

DEMOSTRACIÓN. La demostración se realizará por inducción sobre n. Es evidente que el resultado es cierto para $n = 1$ pues toda matriz de 1×1 es una forma canónica de Jordan. Supóngase que la conclusión es cierta para espacios vectoriales de dimensiones menores que n y que $\dim(V) = n$. Consideraremos dos casos.

CASO 1. rango(T) $< n$. Puesto que R(T) es un subespacio T-invariante de V podemos definir a T_1: $R(T) \to R(T)$ como la restricción de T a R(T). La suposición del Caso 1 nos permite utilizar la hipótesis de inducción a T_1 para concluir que existe una base canónica de Jordan γ para T_1 que

contiene r elementos ($r < n$). Por lo tanto, en virtud del Teorema 6.1(c), γ es una unión disjunta de ciclos de eigenvectores generalizados de T_1 (y por tanto de T). Sean $S_1, S_2, \ldots, S_k$ todos aquellos ciclos en γ que corresponden al eigenvalor cero, y sean y_i y w_i respectivamente los vectores inicial y terminal de S_i. Como $w_i \in \gamma \subseteq R(T)$, existe $x_i \in V$ tal que $T(x_i) = w_i$. Ahora defínase $Y = \{y_1, y_2, \ldots, y_k\}$, $X = \{x_1, x_2, \ldots, x_k\}$ y

$$\gamma_0 = \bigcup_{i=1}^{k} S_i.$$

Supóngase que γ_0 contiene p elementos ($p \leq r$). Recuérdese del Teorema 6.1(a) que cada y_i es un eigenvector que corresponde al eigenvalor cero; por tanto $Y \subseteq N(T)$. Luego, como Y es linealmente independiente (es un subconjunto de γ), puede ser extendido hasta una base $Y \cup Z$ para $N(T)$. Obsérvese que Z debe contener $n - r - k$ elementos puesto que nulidad(T) $= n -$ rango(T) $= n - r$. Sea $S'_i = S_i \cup \{x_i\}$; entonces S'_i es un ciclo de eigenvectores generalizados de T que corresponden al eigenvalor cero y que tiene como vector inicial a y_i. Además, si $z \in Z$, entonces $\{z\}$ es un ciclo correspondiente al eigenvalor cero cuya longitud es 1. Por lo tanto el lema implica que

$$\left(\bigcup_{i=1}^{k} S'_i\right) \cup Z = (\gamma_0 \cup X) \cup Z$$

es un conjunto linealmente independiente que contiene $p + k + (n - r - k) = n - (r - p)$ elementos por el hecho de ser $Y \cup Z$ un conjunto linealmente independiente de vectores iniciales para los ciclos que corresponden al eigenvalor cero.

Demostraremos que $\beta = \gamma \cup X \cup Z$ es la base que se desea encontrar. Primero, obsérvese que si $\gamma_0 = \gamma$ (de manera que $p = r$), entonces $\beta = \gamma_0 \cup X \cup Z$ es un conjunto linealmente independiente que contiene $n - (r - p) = n$ elementos. Entonces β es una base para V. De lo contrario, si $\gamma_0 \neq \gamma$, entonces $\gamma_1 = \{v \in \gamma\colon\ v \notin \gamma_0\}$ es una unión no vacía de ciclos disjuntos de eigenvectores generalizados que corresponden a eigenvalores no nulos $\lambda_2, \ldots, \lambda_m$ de T. Supóngase que

$$0 = \sum_{v \in \beta} a_v v = \sum_{v \in \gamma_0 \cup X \cup Z} a_v v + \sum_{v \in \gamma_1} a_v v.$$

Entonces

$$\sum_{v \in \gamma_0 \cup X \cup Z} (-a_v) v = \sum_{v \in \gamma_1} a_v v.$$

Pero el lado izquierdo es un elemento de K_{λ_1}, donde $\lambda_1 = 0$ y el lado derecho es un elemento de $K_{\lambda_2} + \ldots + K_{\lambda_m}$. Por lo tanto de acuerdo con el Teorema 6.3, ambos lados de la igualdad anterior son iguales a 0.

Así, como $\gamma_0 \cup X \cup Z$ y γ_1 son conjuntos linealmente independientes, se tiene que $a_v = 0$ para toda $v \in \beta$. Por lo tanto β es linealmente independiente, y como $\beta = \gamma \cup X \cup Z$ contiene $r + k + (n - r - k) = n$ elementos, β es una base para V. Pero es evidente que β es una unión disjunta de ciclos de eigenvectores generalizados de T, de modo que β es una base canónica de Jordan para T por el Teorema 6.1.

CASO 2. rango(T) $= n$. Como el polinomio característico de T se descompone como un producto de factores de grado 1, T tiene un eigenvalor λ. Utilícese el Caso 1 para el operador no invertible $\mathsf{T} - \lambda\mathsf{I}$ en V para obtener una base ordenada β para V tal que $[\mathsf{T} - \lambda\mathsf{I}]_\beta = J$ sea una forma canónica de Jordan para $\mathsf{T} - \lambda\mathsf{I}$. Pero entonces $[\mathsf{T}]_\beta = J + \lambda I_n$ es una forma canónica de Jordan para T. ■

Habiendo establecido la existencia de una forma canónica de Jordan, ya podemos obtener varias propiedades importantes de los eigenespacios generalizados.

Teorema 6.5. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V *tal que el polinomio característico de* T *se descomponga en un producto de factores de grado* 1. *Supóngase que* $\lambda_1, \lambda_2, \ldots, \lambda_k$ *son los distintos eigenvalores de* T *y que la multiplicidad de* λ_i *es* $m_i (1 \leq i \leq k)$ *y sea* β *una base canónica de Jordan para* T. *Defínase a* $\beta_i = \beta \cap \mathsf{K}_{\lambda_i} (1 \leq \leq i \leq k)$. *Entonces*

(a) $\mathsf{V} = \mathsf{K}_{\lambda_1} \oplus \mathsf{K}_{\lambda_2} \oplus \cdots \oplus \mathsf{K}_{\lambda_k}$.

(b) β_i *es una base para* K_{λ_i}. *Recíprocamente, si para cada* i λ_i *es una unión de ciclos de eigenvectores generalizados correspondientes a* λ_i *que constituye una base para* K_{λ_i}, *entonces*

$$\bigcup_{i=1}^{k} \gamma_i$$

es una base canónica de Jordan para T.

(c) $\mathsf{K}_{\lambda_i} (1 \leq i \leq k)$ *es un subespacio* T-*invariante de* V.

(d) *Para cada* $i (1 \leq i \leq k)$, $\dim(\mathsf{K}_{\lambda_i}) = m_i$.

(e) *Para cada* $i (1 \leq i \leq k)$, $\mathsf{K}_{\lambda_i} = \mathsf{N}((\mathsf{T} - \lambda_i \mathsf{I})^{m_i})$.

(f) T *es diagonalizable si y sólo si* $\mathsf{E}_{\lambda_i} = \mathsf{K}_{\lambda_i}$ *para cada* $i (1 \leq i \leq \leq k)$.

DEMOSTRACIÓN.

(a) Es evidente que $L(\beta_i) \subseteq \mathsf{K}_{\lambda_i}$. Pero como

$$L\left(\bigcup_{i=1}^{k} \beta_i\right) = L(\beta) = \mathsf{V},$$

se tiene que $\mathsf{K}_{\lambda_1} + \mathsf{K}_{\lambda_2} + \ldots + \mathsf{K}_{\lambda_k} = \mathsf{V}$. El inciso (a) ahora puede deducirse a partir del Teorema 6.3.

(b) Defínase a W_i como el subespacio generado por β_i para $1 \leq i \leq k$. Entonces $W_i \subseteq K_{\lambda_i}$, y por lo tanto $\dim(W_i) \leq \dim(K_{\lambda_i})$. Pero como β es la unión disjunta de $\beta_1, \beta_2, \ldots, \beta_k$ y $L(\beta) = V$, se tiene que $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$. Luego del inciso (a) tenemos que

$$\dim(V) = \sum_{i=1}^{k} \dim(W_i) \leq \sum_{i=1}^{k} \dim(K_{\lambda_i}) = \dim(V).$$

Por tanto $\dim(W_i) = \dim(K_{\lambda_i})$ para $1 \leq i \leq k$. Dado que $W_i \subseteq K_{\lambda_i}$ y $\dim(W_i) = \dim(K_{\lambda_i})$, se tiene que $L(\beta_i) = W_i = K_{\lambda_i}$. Como β_i es linealmente independiente (es un subconjunto de β), β_i es una base para K_{λ_i}.

La recíproca se obtiene del inciso (a) y de los Teoremas 5.13 y 6.1(c).

(c) Recuérdese que K_{λ_i} es un subespacio de V (Teorema 6.2). Ahora bien, β_i es una base para K_{λ_i} formada por ciclos de eigenvectores generalizados que corresponden a λ_i. Pero la imagen bajo T de cualquier vector en un ciclo es claramente una combinación lineal de vectores en dicho ciclo y en consecuencia es un elemento de K_{λ_i}. Luego $T(\beta_i) \subseteq K_{\lambda_i}$, demostrando así que K_{λ_i} es T-invariante.

(d) Defínase a $T_i (1 \leq i \leq k)$ como la restricción de T a K_{λ_i}. Entonces, de acuerdo con el inciso (b), $A_i = [T_i]_{\beta_i}$ es una forma canónica de Jordan para T_i y $[T]_\beta = A_1 \oplus A_2 \oplus \cdots \oplus A_k$. Si $n_i = \dim(K_{\lambda_i})$, entonces el polinomio característico de T_i es $\det(A_i - tI_{n_i}) = (\lambda_i - t)^{n_i}$ puesto que $A_i - tI_{n_i}$ es una matriz triangular superior que tiene a $\lambda_i - t$ en cada posición diagonal. Si $f(t)$ es el polinomio característico de T, entonces

$$\begin{aligned} f(t) &= \det(A_1 - tI_{n_1}) \cdot \det(A_2 - tI_{n_2}) \cdot \ldots \cdot \det(A_k - tI_{n_k}) \\ &= (\lambda_1 - t)^{n_1}(\lambda_2 - t)^{n_2} \ldots (\lambda_k - t)^{n_k}. \end{aligned}$$

Por lo que la multiplicidad de λ_i es n_i; o sea, $m_i = n_i = \dim(K_{\lambda_i})$.

(e) Es evidente que $N((T - \lambda_i I)^{m_i}) \subseteq K_{\lambda_i}$. Supóngase que $x \in K_{\lambda_i}$. Entonces el ciclo S con un vector terminal x, de acuerdo con el Teorema 6.1, es un subconjunto de K_{λ_i} linealmente independiente. Como $\dim(K_{\lambda_i}) = m_i$, se tiene que la longitud de S no puede exceder a m_i; esto es, $(T - \lambda_i I)^p(x) = 0$ para algún entero positivo $p \leq m_i$. Por lo tanto $x \in N((T - \lambda_i I)^{m_i})$, demostrando así que $K_{\lambda_i} \subseteq N((T - \lambda_i I)^{m_i})$.

(f) Si $E_{\lambda_i} = K_{\lambda_i}$ para $1 \leq i \leq k$, entonces, de acuerdo con el inciso (a),

$$E_{\lambda_1} \oplus E_{\lambda_2} \oplus \cdots \oplus E_{\lambda_k} = K_{\lambda_1} \oplus K_{\lambda_2} \oplus \cdots \oplus K_{\lambda_k} = V$$

y así, por el Teorema 5.14, T es diagonalizable.

Recíprocamente, si T es diagonalizable entonces $\dim(E_{\lambda_i}) = m_i$ por el Teorema 5.14. Pero como E_{λ_i} es un subespacio de K_{λ_i} y $\dim(K_{\lambda_i}) = m_i$, por el inciso (d), se tiene que $E_{\lambda_i} = K_{\lambda_i}$ para $1 \leq i \leq k$. ∎

Ejemplo 2. Sea $T: C^3 \to C^3$ definida mediante $T = L_A$, donde

$$A = \begin{pmatrix} 3 & 1 & -2 \\ -1 & 0 & 5 \\ -1 & -1 & 4 \end{pmatrix}.$$

Encontraremos una base para cada eigenespacio y cada eigenespacio generalizado de T.

El polinomio característico de T es

$$f(t) = \det(A - tI) = -(t-3)(t-2)^2.$$

Por lo tanto $\lambda_1 = 3$ y $\lambda_2 = 2$ son los eigenvalores de T con multiplicidades 1 y 2, respectivamente. De acuerdo con el Teorema 6.5, K_{λ_1} tiene dimensión 1, K_{λ_2} tiene dimensión 2 y también $K_{\lambda_1} = N(T - 3I)$ y $K_{\lambda_2} = N((T - 2I)^2)$. Ahora bien $E_{\lambda_1} = N(T - 3I)$ y $E_{\lambda_2} = N(T - 2I)$. Por lo tanto $E_{\lambda_1} = K_{\lambda_1}$. Como

$$A - 3I = \begin{pmatrix} 0 & 1 & -2 \\ -1 & -3 & 5 \\ -1 & -1 & 1 \end{pmatrix}, \qquad \begin{pmatrix} a \\ b \\ c \end{pmatrix} \in E_{\lambda_1} = K_{\lambda_1}$$

si y sólo si

$$\begin{pmatrix} 0 & 1 & -2 \\ -1 & -3 & 5 \\ -1 & -1 & 1 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

o, de manera equivalente, si y sólo si

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

es una solución del sistema

$$\begin{cases} b - 2c = 0 \\ -a - 3b + 5c = 0 \\ -a - b + c = 0. \end{cases}$$

Por el hecho de que el conjunto de soluciones del sistema anterior tenga a

$$\left\{\begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}\right\}$$

como base,

$$\left\{\begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}\right\}$$

es una base para $E_{\lambda_1} = K_{\lambda_1}$.

Análogamente, como

$$A - 2I = \begin{pmatrix} 1 & 1 & -2 \\ -1 & -2 & 5 \\ -1 & -1 & 2 \end{pmatrix}, \qquad \begin{pmatrix} a \\ b \\ c \end{pmatrix} \in \mathsf{E}_{\lambda_2}$$

si y sólo si

$$\begin{pmatrix} 1 & 1 & -2 \\ -1 & -2 & 5 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

o, de manera equivalente, si y sólo si

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

es una solución del sistema

$$\begin{cases} a + b - 2c = 0 \\ -a - 2b + 5c = 0 \\ -a - b + 2c = 0. \end{cases}$$

Una base para el conjunto de soluciones de este sistema, y por tanto para E_{λ_2}, es

$$\left\{ \begin{pmatrix} 1 \\ -3 \\ -1 \end{pmatrix} \right\}.$$

Como

$$(A - 2I)^2 = \begin{pmatrix} 2 & 1 & -1 \\ -4 & -2 & 2 \\ -2 & -1 & 1 \end{pmatrix}, \qquad \begin{pmatrix} a \\ b \\ c \end{pmatrix} \in \mathsf{K}_{\lambda_2}$$

si y sólo si

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

es una solución del sistema

$$\begin{cases} 2a + b - c = 0 \\ -4a - 2b + 2c = 0 \\ -2a - b + c = 0. \end{cases}$$

Una base para el conjunto de soluciones de este sistema, y por tanto para K_{λ_2}, es

$$\left\{ \begin{pmatrix} 1 \\ -3 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} \right\}.$$

Obsérvese que esta base es un ciclo de eigenvectores generalizados correspondientes a λ_2. Por lo tanto de acuerdo con el Teorema 6.5(b)

$$\beta = \left\{ \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -3 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} \right\}$$

es una base para C^3 y

$$[T]_\beta = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

es una forma canónica de Jordan para T.

Ejemplo 3. Sea T: $P_2(C) \to P_2(C)$ definida mediante $T(f) = -f - f'$. De nuevo encontraremos una base para cada eigenespacio y cada eigenespacio generalizado de T. Si $\beta = \{1, x, x^2\}$, entonces β es una base ordenada para $P_2(C)$ y

$$A = [T]_\beta = \begin{pmatrix} -1 & -1 & 0 \\ 0 & -1 & -2 \\ 0 & 0 & -1 \end{pmatrix}.$$

Por lo tanto el polinomio característico de T es $f(t) = \det(A - tI) = -(t+1)^3$. Entonces $\lambda = -1$ es el único eigenvalor de T y por lo tanto, de acuerdo con el Teorema 6.5, $K_\lambda = P_2(C)$ y así cualquier base para $P_2(C)$, por ejemplo β, es una base para K_λ. Ahora bien, $E_\lambda = N(T - \lambda I) = N(T + I)$. Entonces si $f(x) = a + bx + cx^2 \in P_2(C)$, se tendrá que $f(x) \in E_\lambda$ si y sólo si

$$\begin{aligned} 0 &= T(f(x)) + f(x) \\ &= [-(a + bx + cx^2) - (b + 2cx)] + (a + bx + cx^2) \\ &= -(b + 2cx). \end{aligned}$$

Pero $-(b + 2cx) = 0$ si y sólo si $b = c = 0$. Por lo tanto $f(x) \in E_\lambda$ si y sólo si $f(x) = a$ para alguna $a \in C$ y entonces $\{1\}$ es una base para E_λ.

Como $K_\lambda = P_2(C)$, deben existir ciclos de eigenvectores generalizados correspondientes a λ que formen una base canónica de Jordan para T. De hecho, se tiene del Ejercicio 4 que un ciclo único (de longitud 3), y no la unión de dos ciclos (uno de longitud 2 y el otro de longitud 1) ni la unión de tres ciclos (todos de longitud 1), formará una base para $P_2(C)$. Tal ciclo es $\gamma = \{2, -2x, x^2\}$ y

$$[T]_\gamma = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & -1 \end{pmatrix}$$

es una forma canónica de Jordan para T. Veremos en la sección siguiente cómo encontrar tal base canónica de Jordan.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

   (a) Los eigenvectores de un operador lineal T son también eigenvectores generalizados de T.
   (b) Es posible que un eigenvector generalizado de un operador lineal T esté asociado con un escalar que no sea un eigenvalor de T.
   (c) Cualquier operador lineal en un espacio vectorial dimensionalmente finito tiene una forma canónica de Jordan.
   (d) Los ciclos de eigenvectores generalizados son linealmente independientes.
   (e) Existe exactamente un ciclo de eigenvectores generalizados correspondientes a cada eigenvalor de un operador lineal en un espacio vectorial dimensionalmente finito.
   (f) Sea T un operador lineal en un espacio vectorial dimensionalmente finito cuyo polinomio característico se descompone en un producto de factores de grado 1, y sean $\lambda_1, \lambda_2, \ldots, \lambda_k$ los distintos eigenvalores de T. Si, para cada i, β_i es una base cualquiera para K_{λ_i}, entonces $\beta_1 \cup \beta_2 \cup \ldots \cup \beta_k$ es una base canónica de Jordan para T.
   (g) Para cualquier bloque de Jordan J, L_J tiene una forma canónica de Jordan J.
   (h) Sea T un operador lineal en un espacio vectorial n-dimensional cuyo polinomio característico se descompone en un producto de factores de grado 1. Para cualquier eigenvalor λ de T, $K_\lambda = N((T - \lambda I)^n)$.

2. Para cada uno de los siguientes operadores lineales T, encontrar una base para cada eigenespacio y para cada eigenespacio generalizado.

   (a) $T = L_A$, donde

$$A = \begin{pmatrix} 1 & 1 \\ -1 & 3 \end{pmatrix}$$

   (b) $T = L_A$, donde

$$A = \begin{pmatrix} 11 & -4 & -5 \\ 21 & -8 & -11 \\ 3 & -1 & 0 \end{pmatrix}$$

   (c) $T\colon\ P_2(C) \to P_2(C)$ definido mediante $T(f) = 2f - f'$

3.* Sea S un ciclo de eigenvectores generalizados de un operador lineal T en V que corresponde al eigenvalor λ. Demostrar que el subespacio generado por S es un subespacio T-invariante de V.

4. Sea β una base canónica de Jordan para un operador lineal T en V y sea $[T]_\beta = J_1 \oplus J_2 \oplus \cdots \oplus J_k$, donde cada $J_i (1 \leq i \leq k)$ es un bloque de Jordan. Sea λ un eigenvalor de T y sea m el número de bloques de Jordan que tienen a λ en cada una de las posiciones de la diagonal. Demostrar que $1 \leq m \leq \dim(E_\lambda)$. Veremos más adelante que $m = \dim(E_\lambda)$.

5. Sea T: $V \to W$ una transformación lineal. Demostrar los incisos siguientes:

   (a) $N(T) = N(-T)$.
   (b) $N(T^k) = N((-T)^k)$ para cualquier entero positivo k.
   (c) Si $W = V$ (tal que T es un operador lineal en V) y λ es un eigenvalor de T, entonces para cualquier entero positivo k

$$N((T - \lambda I_V)^k) = N((\lambda I_V - T)^k).$$

6. Sea U un operador lineal en un espacio vectorial dimensionalmente finito V. Demostrar los siguientes incisos:

   (a) $N(U) \subseteq N(U^2) \subseteq \ldots \subseteq N(U^k) \subseteq N(U^{k+1}) \subseteq \ldots$.
   (b) Si $\text{rango}(U^m) = \text{rango}(U^{m+1})$ para algún entero positivo m, entonces $\text{rango}(U^m) = \text{rango}(U^k)$ para cualquier entero positivo $k \geq m$.
   (c) Si $\text{rango}(U^m) = \text{rango}(U^{m+1})$ para algún entero positivo m, entonces $N(U^m) = N(U^k)$ para cualquier entero positivo $k \geq m$.
   (d) Sea T un operador lineal y sea λ un eigenvalor de T. Demostrar que si $\text{rango}((T - \lambda I)^m) = \text{rango}((T - \lambda I)^{m+1})$ para algún entero m, entonces $K_\lambda = N((T - \lambda I)^m)$.
   (e) *Segunda prueba para diagonalizabilidad.* Sea T un operador lineal cuyo polinomio característico se descompone en un producto de factores de grado 1. Supóngase que $\lambda_1, \lambda_2, \ldots, \lambda_k$ son los distintos eigenvalores de T. Entonces T es diagonalizable si y sólo si $\text{rango}(T - \lambda_i I) = \text{rango}((T - \lambda_i I)^2)$ para $1 \leq i \leq k$.
   (f) Utilizar el inciso (e) para obtener una demostración más sencilla del Ejercicio 10(d) de la Sección 5.4: Si T es un operador lineal diagonalizable en un espacio vectorial dimensionalmente finito V y W es cualquier subespacio T-invariante de V, entonces T_W es diagonalizable.

7. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V tal que el polinomio característico $f(t)$ de T se descompone en un producto de factores de grado 1. Demostrar que $f(T) = T_0$; esto es, demostrar que T satisface a su polinomio característico. (Este es un caso especial del Teorema de Cayley-Hamilton.) *Sugerencia:* Demostrar que si β es una base canónica de Jordan para T, entonces $f(T)(x) = 0$ para cada $x \in \beta$.

8. Este ejercicio tiene como finalidad ilustrar la demostración del Caso 1 del Teorema 6.4 para una transformación lineal particular T: $C^n \to C^n$. (Ver también el Ejemplo 1.)

Sea $T: C^{11} \to C^{11}$ definida mediante

$$T(u) = (a_1 + 2a_2 - a_3, \ -a_1 - 5a_2 + 3a_3, \ -2a_1 - 7a_2 + 4a_3, \\ 6a_4 - 9a_5, 4a_4 - 6a_5, a_6 + a_7, -a_6 + 3a_7, 3a_8, 0, 0, 0),$$

donde $u = (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11})$.

(a) Sea $\gamma = \{y_1, w_1, y_2, v_1, v_2, v_3\}$, donde $y_1 = e_1 - 2e_2 - 3e_3$, $w_1 = 2e_1 - 3e_2 - 5e_3$, $y_2 = 3e_4 + 2e_5$, $v_1 = e_6 + e_7$, $v_2 = 3e_6 + 4e_7$ y $v_3 = e_3$. Demostrar que γ es una base ordenada para $R(T)$. *Sugerencia:* Tomando a u como en el párrafo anterior

$$T(u) = (-a_1 + 4a_2 - 3a_3)y_1 + (a_1 - a_2 + a_3)w_1 + (2a_4 - 3a_5)y_2 \\ + (7a_6 - 5a_7)v_1 + (-2a_6 + 2a_7)v_2 + (3a_8)v_3.$$

(b) Dedúzcase que r, el rango de T, es igual a 6 y que la nulidad de T es 5.

(c) Sea T_1 la restricción de T a $R(T)$. Demostrar que γ es una base canónica de Jordan para T_1.

(d) Demostrar que $S_1 = \{y_1, w_1\}$, $S_2 = \{y_2\}$, $S_3 = \{v_1, v_2\}$ y $S_4 = \{v_3\}$ son ciclos de eigenvectores generalizados de T_1 correspondientes respectivamente a los eigenvalores 0, 0, 2 y 3. (Por lo cual, en la notación de la demostración del Teorema 6.4, $k = 2$, $\gamma_0 = \{y_1, w_1, y_2\}$ y $p = 3$.)

(e) Sean $x_1 = -e_1 + 5e_2 + 7e_3$ y $x_2 = 8e_4 + 5e_5$. Demostrar que $T(x_i) = w_i$ para $i = 1, 2$. Hágase $X = \{x_1, x_2\}$.

(f) Obsérvese que en la notación de la demostración del Teorema 6.4 los vectores y_2 y w_2 son iguales, y hágase $Y = \{y_1, y_2\}$. Defínanse $z_1 = e_9$, $z_2 = e_{10}$ y $z_3 = e_{11}$. Demuéstrese que $Z = \{z_1, z_2, z_3\}$ es un conjunto (que contiene $n - r - k$ elementos) tal que $Y \cup Z$ es una base para $N(T)$.

(g) Defínanse $S'_1 = \{y_1, w_1, x_1\}$ y $S'_2 = \{w_2, x_2\}$. Demostrar que S'_1 y S'_2 son ciclos de eigenvectores generalizados de T correspondientes al eigenvalor cero. Entonces, de acuerdo con el lema del Teorema 6.4,

$$\left(\bigcup_{i=1}^{k} S'_i\right) \cup Z = (\gamma_0 \cup X) \cup Z$$

es un conjunto linealmente independiente que contiene a $n - (r - p) = 8$ elementos.

(h) La demostración del Teorema 6.4 muestra que $\beta = \gamma \cup X \cup Z$ es un conjunto linealmente independiente. Considerando este hecho dedúzcase que β es una base para C^{11}.

(i) Finalmente, demostrar que β es una base canónica de Jordan para T mediante el cálculo de $[T]_\beta$.

9. Demostrar los incisos (a) y (c) del Teorema 6.1.

10. Sea T un operador lineal en un espacio vectorial dimensionalmente finito. Supóngase que $\lambda_1, \lambda_2, \ldots, \lambda_k$ son los eigenvalores distintos de T y que el mayor bloque de Jordan correspondiente a λ_j en una forma canónica de Jordan de T es de tamaño $p_j \times p_j$. Demostrar que el polinomio mínimo de T es

$$(t - \lambda_1)^{p_1}(t - \lambda_2)^{p_2} \ldots (t - \lambda_k)^{p_k}.$$

6.2 FORMA CANONICA DE JORDAN

Para los fines de esta sección fijaremos un operador lineal T en un espacio vectorial n-dimensional V tal que el polinomio característico de T se descompone en un producto de factores de grado 1. Sean $\lambda_1, \lambda_2, \ldots, \lambda_k$ los distintos eigenvalores de T.

El Teorema 6.4 asegura la existencia de una base canónica de Jordan β para T; esto es, $J = [\mathsf{T}]_\beta$ es una forma canónica de Jordan para T. Resumamos brevemente los resultados de la Sección 6.1. Para cada $i = 1, 2, \ldots k$ existe una base β_i para K_{λ_i} tal que β_i es una unión disjunta de ciclos correspondientes al eigenvalor λ_i y

$$\beta = \bigcup_{i=1}^{k} \beta_i.$$

Sea T_i la restricción de T a K_{λ_i}. Entonces $A_i = [\mathsf{T}_i]_{\beta_i}$ es una forma canónica de Jordan para T_i y

$$J = [\mathsf{T}]_\beta = A_1 \oplus A_2 \oplus \cdots \oplus A_k$$

es una forma canónica de Jordan para T.

En esta sección calcularemos las matrices A_i y las bases β_i calculando también así J y β. Mientras se desarrolla un método para encontrar a J, se hará evidente que en cierto sentido las matrices A_i son únicas. Lo que queremos decir por "en cierto sentido" se hará más claro a medida que avancemos.

Para ayudarnos en la formulación de un teorema de unicidad para J adoptaremos la siguiente convención: La base β_i para K_{λ_i} se ordenará en adelante de tal modo que los ciclos aparecerán en orden de longitud decreciente. Esto es, si β_i es una unión disjunta de ciclos $S_1, S_2, \ldots, S_{k_i}$ y si la longitud del ciclo S_j es p_j, pondremos índices a los ciclos de modo que $p_1 \geq p_2 \geq \ldots \geq p_{k_i}$. Esta ordenación de los ciclos determina una ordenación para β_i y por lo tanto determina a la matriz A_i. Es en este sentido que la matriz A_i es única. Se deduce entonces que la forma canónica de Jordan para T es única para un ordenamiento de los eigenvalores de T. Como también veremos, no existe ningún teorema de unicidad comparable para las bases β_i o para β. Específicamente, lo que se demostrará es que el número k_i de ciclos que forman a β_i y la longitud $p_j (j = 1, 2, \ldots, k_i)$ de cada ciclo está completamente determinada por T.

Ejemplo 4. Para ilustrar que la matriz A_i queda totalmente determinada por los números k_1, p_1, $p_2, \ldots, p_{k_i}$ supóngase que $k_i = 4$ (esto es, existen cuatro ciclos), $p_1 = 3$, $p_2 = 3$, $p_3 = 2$ y $p_4 = 1$. Entonces

$$A_i = \begin{pmatrix} \lambda_i & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_i & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_i & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_i & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_i & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_i & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_i & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_i \end{pmatrix};$$

esto es, A_i es una suma directa de la forma $J_1 \oplus J_2 \oplus J_3 \oplus J_4$.

Como una ayuda para calcular A_i y β_i introduciremos un arreglo de puntos, llamado *diagrama de puntos,* para ayudarnos a visualizar la forma de la matriz A_i y de la base β_i. Supóngase, como antes, que β_i es una unión disjunta de ciclos S_1, S_2, ... , S_{k_i} con longitudes respectivas $p_1 \geq p_2 \geq \ldots \geq p_{k_i}$. El diagrama de puntos contiene un punto para cada miembro de β_i y se construye de acuerdo con las reglas siguientes.

1. El arreglo consta de k_i columnas (una columna para cada ciclo).
2. Contando de izquierda a derecha, la columna j consta de p_j puntos que corresponden a los miembros de S_j de la siguiente manera: Si x_j es el vector terminal de S_j entonces el punto de arriba corresponde a $(\mathsf{T} - \lambda_i \mathsf{I})^{p_j-1}(x_j)$; el segundo punto, a $(\mathsf{T} - \lambda_i \mathsf{I})^{p_j-2}(x_j)$; etc. Por lo tanto, el punto final (el de más abajo) de la columna corresponde a x_j.

Así, el diagrama asociado con β_i puede ser descrito como

$$\begin{array}{llll}
\cdot(\mathsf{T} - \lambda_i\mathsf{I})^{p_1-1}(x_1) & \cdot(\mathsf{T} - \lambda_i\mathsf{I})^{p_2-1}(x_2) & \cdots & \cdot(\mathsf{T} - \lambda_i\mathsf{I})^{p_{k_i}-1}(x_{k_i}) \\
\cdot(\mathsf{T} - \lambda_i\mathsf{I})^{p_1-2}(x_1) & \cdot(\mathsf{T} - \lambda_i\mathsf{I})^{p_2-2}(x_2) & & \cdot(\mathsf{T} - \lambda_i\mathsf{I})^{p_{k_i}-2}(x_{k_i}) \\
\quad\cdot & \quad\cdot & & \quad\cdot \\
 & \quad\cdot & & \quad\cdot \\
\quad\cdot & & & \quad\cdot \\
 & \quad\cdot & & \cdot(\mathsf{T} - \lambda_i\mathsf{I})(x_{k_i}) \\
\quad\cdot & \cdot(\mathsf{T} - \lambda_i\mathsf{I})(x_2) & & \cdot x_{k_i} \\
\cdot(\mathsf{T} - \lambda_i\mathsf{I})(x_1) & \cdot x_2 & & \\
\cdot x_1 & & &
\end{array}$$

En el diagrama anterior hemos identificado cada punto con el miembro de β_i al que corresponde.

Nótese que el diagrama de puntos para β_i tiene k_i columnas (una para cada ciclo) y p_1 renglones. Obsérvese también que como $p_1 \geq p_2 \geq \geq \ldots \geq p_{k_i}$, las columnas del diagrama de puntos se hacen más cortas (o al menos no más largas) a medida que nos movemos de izquierda a derecha.

Puede observarse también que si r_j es el número de puntos en la columna j del arreglo, entonces $r_1 \geq r_2 \geq \ldots \geq r_{p_1}$. Como la demostración de este hecho es de naturaleza combinatoria, la dejaremos para los ejercicios.

Regresando al Ejemplo 4, donde $k_i = 4$, $p_1 = 3$, $p_2 = 3$, $p_3 = 2$ y $p_4 = 1$, vemos que el diagrama de puntos para β_i es

$$
\begin{array}{cccc}
\bullet & \bullet & \bullet & \bullet \\
\bullet & \bullet & \bullet & \\
\bullet & \bullet & &
\end{array}
$$

Obtendremos un método para calcular el diagrama de puntos para β_i únicamente en términos de T, por lo que el diagrama de puntos queda determinado en forma única por T. Es importante, sin embargo, entender que cuando decimos que el diagrama de puntos queda únicamente determinado por T no estamos haciendo aseveraciones sobre la unicidad de β_i. De hecho, como veremos, la base β_i no es única. Por la unicidad del diagrama de puntos entendemos que si β_i y β'_i son dos bases canónicas de Jordan para K_{λ_i}, entonces los diagramas de puntos para β_i y β'_i son idénticos. Luego, si β'_i es una unión disjunta de k'_i ciclos de longitudes $p'_1 \geq p'_2 \geq \cdots \geq p'_{k_i}$, entonces $k'_i = k_i$ y $p'_1 = p_1, p'_2 = p_2, \ldots, p'_{k_i} = p_{k_i}$.

Para establecer este resultado de unicidad, utilizaremos el siguiente hecho de carácter combinatorio: Cualquier diagrama de puntos queda completamente determinado por el número de sus renglones y por el número de puntos en cada renglón. (Véase Ejercicio 7.) Así, si estos números pudieran obtenerse a partir de las propiedades intrínsecas de la transformación T (por ejemplo, como los rangos de $(T - \lambda_i I)^j$ para distintos valores de j), se podría construir el diagrama de puntos y se demostraría la unicidad de los números k_i, p_1, $p_2, \ldots, p_{k_i}$. Los resultados siguientes proporcionan el método deseado para calcular dichos números.

Teorema 6.6. *Para cualquier entero positivo* r *los vectores de la base en* β_i *que estén asociados con puntos en los primeros* r *renglones de un diagrama de puntos para* β_i *forman una base para* $N((T - \lambda_i I)^r)$. *Por lo tanto, el número de puntos en los primeros* r *renglones de un diagrama de puntos para* β_i *es la nulidad de* $(T - \lambda_i I)^r$.

DEMOSTRACIÓN. Los vectores de la base en β_i que están asociados con puntos en los primeros r renglones de un diagrama de puntos para β_i son los primeros r elementos de los ciclos $S_j (j = 1, 2, \ldots, k_i)$ que forman a β_i. Por lo tanto, estos vectores de la base son elementos de $N((T - \lambda_i I)^r)$.

Además, estos vectores son linealmente independientes puesto que forman un subconjunto de β_i, de manera que basta con demostrar que estos vectores de la base generan a $\mathsf{N}((\mathsf{T}-\lambda_i\mathsf{I})^r)$.

Para cada $j(j = 1, 2, \ldots, k_i)$, sea $\mathsf{W}_j = L(S_j)$. Como W_j es T-invariante por el Ejercicio 3 de la Sección 6.1, es también $(\mathsf{T}-\lambda_i\mathsf{I})^r$-invariante. Además, de acuerdo con el Teorema 5.13, $\mathsf{K}_{\lambda_i} = \mathsf{W}_1 \oplus \mathsf{W}_2 \oplus \cdots \oplus \mathsf{W}_{k_i}$ Si $x \in \mathsf{N}((\mathsf{T}-\lambda_i\mathsf{I})^r)$, entonces por definición $x \in \mathsf{K}_{\lambda_i}$. Luego existen elementos únicos $w_j \in \mathsf{W}_j (j = 1, 2, \ldots, k_i)$ tales que $x = w_1 + w_2 + \ldots + w_{k_i}$. Por lo tanto

$$\begin{aligned} 0 &= (\mathsf{T}-\lambda_i\mathsf{I})^r(x) \\ &= (\mathsf{T}-\lambda_i\mathsf{I})^r(w_1) + (\mathsf{T}-\lambda_i\mathsf{I})^r(w_2) + \cdots + (\mathsf{T}-\lambda_i\mathsf{I})^r(w_{k_i}). \end{aligned}$$

Se tiene que

$$(\mathsf{T}-\lambda_i\mathsf{I})^r(w_j) = 0 \qquad \text{para } j = 1, 2, \ldots, k_i.$$

Supóngase para cada j que

$$S_j = \{(\mathsf{T}-\lambda_i\mathsf{I})^{p_j-1}(x_j), (\mathsf{T}-\lambda_i\mathsf{I})^{p_j-2}(x_j), \ldots, (\mathsf{T}-\lambda_i\mathsf{I})(x_j), x_j\}.$$

Entonces como

$$w_j = a_{p_j-1}(\mathsf{T}-\lambda_i\mathsf{I})^{p_j-1}(x_j) + \cdots + a_1(\mathsf{T}-\lambda_i\mathsf{I})(x_j) + a_0 x_j$$

Para algunos escalares $a_{p_j-1}, \ldots, a_1, a_0$,

$$0 = (\mathsf{T}-\lambda_i\mathsf{I})^r(w_j) = a_{p_j-r-1}(\mathsf{T}-\lambda_i\mathsf{I})^{p_j-1}(x_j) + \cdots + a_0(\mathsf{T}-\lambda_i\mathsf{I})^r(x_j).$$

Dado que S_j es linealmente independiente, se tiene que $a_{p_j-r-1} = \ldots = a_0 = 0$. Por lo tanto

$$\begin{aligned} w_j = a_{p_j-1}&(\mathsf{T}-\lambda_i\mathsf{I})^{p_j-1}(x_j) + a_{p_j-2}(\mathsf{T}-\lambda_i\mathsf{I})^{p_j-2}(x_j) \\ &+ \cdots + a_{p_j-r}(\mathsf{T}-\lambda_i\mathsf{I})^{p_j-r}(x_j). \end{aligned}$$

Esto es que w_j es una combinación lineal de los vectores de la base en β_i que están asociados con puntos en los primeros r renglones de la columna j de un diagrama de puntos para β_i, y entonces $x = w_1 + w_2 + \ldots + w_{k_i}$ es una combinación lineal de miembros de β_i asociados con puntos en los primeros r renglones de un diagrama de puntos para β_i. Concluimos que estos vectores forman una base para $\mathsf{N}((\mathsf{T}-\lambda_i\mathsf{I})^r)$. ■

En el caso en que $r = 1$, el Teorema 6.6 da origen al corolario siguiente.

Corolario. *Sea β_i una base canónica de Jordan para la restricción de* T *a* K_{λ_i} *y supóngase que β_i es la unión disjunta de* k_i *ciclos de eigenvectores generalizados correspondientes a* λ_i. *Entonces la dimensión de* E_{λ_i} *es igual a* k_i. *Por lo tanto, en una forma canónica de Jordan para* T, *el número de*

bloques de Jordan correspondientes al eigenvalor λ_i es igual a la dimensión de E_{λ_i}.

Ahora somos capaces de formular un procedimiento para calcular el diagrama de puntos para β_i directamente a partir de T.

Teorema 6.7. *Sea r_j el número de puntos en el renglón j de un diagrama de puntos para β_i. Entonces*

(a) $r_1 = dim(V) - rango(T - \lambda_i I)$.

(b) $r_j = rango((T - \lambda_i I)^{j-1}) - rango((T - \lambda_i I)^j)$ *si* $j > 1$.

DEMOSTRACIÓN. De acuerdo con el Teorema 6.6,

$$\begin{aligned} r_1 + r_2 + \ldots + r_j &= \text{nulidad}((T - \lambda_i I)^j) \\ &= \dim(V) - \text{rango}((T - \lambda_i I)^j) \text{ para cualquier } j \geq 1. \end{aligned}$$

Por lo tanto

$$r_1 = \dim(V) - \text{rango}((T - \lambda_i I)^1)$$

y

$$\begin{aligned} r_j &= (r_1 + r_2 + \ldots + r_j) - (r_1 + r_2 + \ldots + r_{j-1}) \\ &= (\dim(V) - \text{rango}((T - \lambda_i I)^j)) - (\dim(V) - \text{rango}((T - \lambda_i I)^{j-1})) \\ &= \text{rango}((T - \lambda_i I)^{j-1}) - \text{rango}((T - \lambda_i I)^j) \text{ para } j > 1. \quad \blacksquare \end{aligned}$$

Este teorema muestra que un diagrama de puntos para β_i está completamente determinado por T. Por lo tanto hemos demostrado el siguiente resultado de unicidad.

Corolario. *Para cualquier eigenvalor λ_i de T el diagrama de puntos para β_i es único. Por lo tanto, sujeta a la convención de que los ciclos se encuentran en orden de longitud decreciente, la forma canónica de Jordan de un operador lineal es única hasta el ordenamiento de sus eigenvalores.*

Antes de dar algunos ejemplos del uso del Teorema 6.7 definiremos de la manera evidente la forma canónica de Jordan para una matriz.

Definición. *Sea A una matriz de $n \times n$ con elementos de F tal que el polinomio característico de A (y por lo tanto de L_A) se descompone en un producto de factores de grado 1. Entonces la* forma canónica de Jordan *de A se define como la forma canónica de Jordan del operador lineal L_A en F^n.*

Obsérvese que si J es la forma canónica de Jordan de una matriz A, entonces J y A son similares. De hecho, si $\beta = \{z_1, z_2, \ldots, z_n\}$ es una base canónica de Jordan para L_A y Q es la matriz de $n \times n$ que tiene a z_j como su columna j, entonces $J = Q^{-1}AQ$ en virtud del Teorema 5.1.

En los tres ejemplos siguientes calcularemos la forma canónica de Jordan para dos matrices y un operador lineal.

Ejemplo 5. Sea

$$A = \begin{pmatrix} 2 & -1 & 0 & 1 \\ 0 & 3 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & -1 & 0 & 3 \end{pmatrix}.$$

Encontraremos la forma canónica de Jordan de A y una base canónica de Jordan para la transformación lineal L_A. El polinomio característico de A es

$$\det(A - tI) = (t-2)^3(t-3).$$

Luego, A tiene dos eigenvalores distintos, $\lambda_1 = 2$ y $\lambda_2 = 3$ con multiplicidades respectivas 3 y 1.

Sea β_1 una base canónica de Jordan para la restricción de L_A para K_{λ_1}. Como λ_1 tiene multiplicidad 3, $\dim(\mathsf{K}_{\lambda_1}) = 3$ en virtud del Teorema 6.5. De modo que el diagrama de puntos para β_1 contiene tres puntos. Como antes, sea r_j el número de puntos en el renglón j del diagrama. Aplicando el Teorema 6.7 tenemos

$$r_1 = 4 - \text{rango}(A - 2I) = 4 - \text{rango}\begin{pmatrix} 0 & -1 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} = 4 - 2 = 2$$

y

$$r_2 = \text{rango}(A - 2I) - \text{rango}((A - 2I)^2) = 2 - 1 = 1.$$

(En realidad, en este caso no es necesario calcular a r_2. Hubiéramos podido deducir que $r_2 = 1$ del hecho de que $r_1 = 2$ y que el diagrama tiene tres puntos.) Por lo tanto, el diagrama de puntos asociado con β_1 es

$$\begin{matrix} \bullet & \bullet \\ \bullet & \end{matrix}$$

Así, si T_i es la restricción de L_A a $\mathsf{K}_{\lambda_i} (i = 1, 2)$, debemos tener que

$$A_1 = [\mathsf{T}_1]_{\beta_1} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

Como $\dim(\mathsf{K}_{\lambda_2}) = 1$, cualquier β_2 para K_{λ_2} estará formada de un eigenvector único que corresponde a $\lambda_2 = 3$. Entonces

$$A_2 = [\mathsf{T}_2]_{\beta_2} = (3).$$

Haciendo $\beta = \beta_1 \cup \beta_2$, tenemos que

$$J = [\mathsf{L}_A]_\beta = A_1 \oplus A_2 = \left(\begin{array}{ccc|c} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ \hline 0 & 0 & 0 & 3 \end{array}\right),$$

y por lo tanto J es la forma canónica de Jordan para A.

Busquemos ahora una base canónica de Jordan para $\mathsf{T} = \mathsf{L}_A$, para lo cual debemos encontrar primero una base canónica de Jordan β_1 para T_1. Sabemos de los cálculos anteriores que el diagrama de puntos correspondiente a β_1 debe ser

$$\begin{array}{ll} \bullet(\mathsf{T} - \lambda_i \mathsf{I})(x_1) & \bullet x_2 \\ \bullet x_1 & \end{array}$$

De este diagrama vemos que debemos seleccionar a x_1 tal que $x_1 \in \mathsf{N}((\mathsf{T} - \lambda_1 \mathsf{I})^2)$ pero que $x_1 \notin \mathsf{N}((\mathsf{T} - \lambda_1 \mathsf{I})^1)$. Como

$$A - 2I = \begin{pmatrix} 0 & -1 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}, \qquad (A - 2I)^2 = \begin{pmatrix} 0 & -2 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -2 & 1 & 1 \end{pmatrix}.$$

Se puede ver ahora fácilmente que

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 2 \end{pmatrix} \right\}$$

es una base para $\mathsf{N}((\mathsf{T} - \lambda_1 \mathsf{I})^2) = \mathsf{K}_{\lambda_1}$. De estos vectores básicos,

$$\begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} 0 \\ 1 \\ 0 \\ 2 \end{pmatrix}$$

satisfacen la condición de no pertenecer a $\mathsf{N}((\mathsf{T} - \lambda_1 \mathsf{I})^1)$. Por lo tanto podemos seleccionar a x_1 de manera que sea cualquiera de los dos vectores. Tomaremos

$$x_1 = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \end{pmatrix}.$$

Entonces

$$(\mathsf{T} - \lambda_1\mathsf{I})(x_1) = (A - 2I)(x_1) = \begin{pmatrix} 0 & -1 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ -1 \\ -1 \end{pmatrix}.$$

Ahora simplemente tómese a x_2 como un elemento de E_{λ_1} que sea linealmente independiente de

$$(\mathsf{T} - \lambda_1\mathsf{I})(x_1) = \begin{pmatrix} -1 \\ -1 \\ -1 \\ -1 \end{pmatrix};$$

por ejemplo, selecciónese

$$x_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Así, hemos asociado la base canónica de Jordan

$$\beta_1 = \left\{\begin{pmatrix} -1 \\ -1 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}\right\}$$

con el diagrama de puntos de la siguiente manera:

$$\begin{matrix} \cdot\begin{pmatrix} -1 \\ -1 \\ -1 \\ -1 \end{pmatrix} & \cdot\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \\ \cdot\begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \end{pmatrix} & \end{matrix}$$

Al lector podría preocuparle que no se haya verificado la independencia lineal de β_1. Sin embargo, debe estar seguro de que esta verificación no es necesaria en virtud del lema del Teorema 6.4. Dado que se seleccionó a x_2 tal que fuera linealmente independiente del vector inicial $(\mathsf{T} - \lambda_1\mathsf{I})(x_1)$ del ciclo $\{(\mathsf{T} - \lambda_1\mathsf{I})(x_1), x_1\}$, se deduce de este lema que β_1 es linealmente independiente.

Cualquier eigenvector de L_A que corresponda al eigenvalor $\lambda_2 = 3$ formará la base deseada β_2 para K_{λ_2} —por ejemplo,

$$\beta_2 = \left\{\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}\right\}.$$

Así,

$$\beta = \beta_1 \cup \beta_2 = \left\{\begin{pmatrix} -1 \\ -1 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}\right\}$$

es una base canónica de Jordan para L_A.

Nótese que si

$$Q = \begin{pmatrix} -1 & 0 & 1 & 1 \\ -1 & 1 & 0 & 0 \\ -1 & 2 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix},$$

entonces $J = Q^{-1}AQ$.

Ejemplo 6. Sea

$$A = \begin{pmatrix} 2 & -4 & 2 & 2 \\ -2 & 0 & 1 & 3 \\ -2 & -2 & 3 & 3 \\ -2 & -6 & 3 & 7 \end{pmatrix}.$$

De nuevo encontraremos una forma canónica de Jordan de J para A y una matriz Q tal que $J = Q^{-1}AQ$.

El polinomio característico de A es $\det(A - tI) = (t-2)^2(t-4)^2$. Sean $\lambda_1 = 2$, $\lambda_2 = 4$ y β_i la base canónica de Jordan para T_i, la restricción de L_A *a* K_{λ_i}, para $i = 1, 2$.

Principiamos calculando el diagrama de puntos para β_1. Sea r_1 el número de puntos en el primer renglón de este diagrama; entonces $r_1 = 4 - \text{rango}(A - 2I) = 4 - 2 = 2$, de modo que el diagrama de puntos para β_1 es

$$\cdot\ \cdot$$

Entonces

$$A_1 = [T_1]_{\beta_1} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Calcularemos ahora el diagrama de puntos para β_2. Como $\text{rango}(A - 4I) = 3$, existe únicamente $4 - 3 = 1$ punto en el primer renglón del

diagrama. Como K_{λ_2} tiene dimensión 2 (Teorema 6.5), el diagrama de puntos para β_2 debe ser

$$\begin{matrix}\bullet\\ \bullet\end{matrix}$$

Luego

$$A_2 = [T_2]_{\beta_2} = \begin{pmatrix} 4 & 1 \\ 0 & 4 \end{pmatrix}.$$

Entonces si $\beta = \beta_1 \cup \beta_2$, la forma canónica de Jordan de L_A es

$$J = [L_A]_\beta = \left(\begin{array}{cc|cc} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ \hline 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 4 \end{array}\right).$$

Con el objeto de encontrar una matriz Q tal que $Q^{-1}AQ = J$, primero debemos encontrar una base canónica de Jordan β para T. El diagrama de puntos para β_1 indica que β_1 se puede escoger como cualquier conjunto linealmente independiente de eigenvalores de A correspondientes a $\lambda_1 = 2$. Por ejemplo,

$$\beta_1 = \left\{ \begin{pmatrix} 2 \\ 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \end{pmatrix} \right\}$$

será suficiente. Para β_2 debemos encontrar un elemento $x_1 \in K_{\lambda_2} = N((L_A - \lambda_2 I)^2)$ tal que $x_1 \notin N((L_A - \lambda_2 I)^1)$. Una manera para encontrar dicho elemento fue utilizada en el Ejemplo 5 para seleccionar al vector x_1. En este ejemplo ilustraremos otro método para obtener tal vector. Un cálculo sencillo muestra que una base para el espacio nulo de $L_A - \lambda_2 I$ es

$$\left\{ \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

Sea

$$(A - 4I)(x_1) = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

y tómese a x_1 tal que sea la preimagen de

$$\begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Para realizar esta operación debemos encontrar una solución a la ecuación matricial

$$(A - 4I)\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix};$$

esto es

$$\begin{pmatrix} -2 & -4 & 2 & 2 \\ -2 & -4 & 1 & 3 \\ -2 & -2 & -1 & 3 \\ -2 & -6 & 3 & 3 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

Puede verificarse fácilmente que

$$\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 0 \end{pmatrix}$$

es una solución; de esta manera tenemos

$$x_1 = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 0 \end{pmatrix}.$$

Entonces

$$\beta_2 = \{(\mathsf{L}_A - \lambda_2\mathsf{I})(x_1), x_1\} = \left\{\begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ -1 \\ 0 \end{pmatrix}\right\}.$$

Por lo tanto

$$\beta = \beta_1 \cup \beta_2 = \left\{\begin{pmatrix} 2 \\ 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ -1 \\ 0 \end{pmatrix}\right\}$$

es una base canónica de Jordan para L_A.

Así, si

$$Q = \begin{pmatrix} 2 & 0 & 0 & 1 \\ 1 & 1 & 1 & -1 \\ 0 & 2 & 1 & -1 \\ 2 & 0 & 1 & 0 \end{pmatrix},$$

entonces $J = Q^{-1}AQ$.

Ejemplo 7. Sea V el espacio vectorial de las funciones polinómicas sobre R en dos variables x e y de grado a lo más 2. (Una base para V es $a = \{1, x, y, x^2, y^2, xy\}$.) Considérese el mapeo T: V → V definido por

$$\mathsf{T}(f) = \frac{\partial}{\partial x} f.$$

Por ejemplo, si $f(x, y) = x + 2x^2 - 3xy + y$, entonces

$$\mathsf{T}(f) = \frac{\partial}{\partial x} f(x, y) = 1 + 4x - 3y.$$

Encontraremos una base canónica de Jordan para T.

Primero, obsérvese que si $A = [\mathsf{T}]_\alpha$, entonces

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Así, el polinomio característico de T es

$$\det(A - tI) = \det \begin{pmatrix} -t & 1 & 0 & 0 & 0 & 0 \\ 0 & -t & 0 & 2 & 0 & 0 \\ 0 & 0 & -t & 0 & 0 & 1 \\ 0 & 0 & 0 & -t & 0 & 0 \\ 0 & 0 & 0 & 0 & -t & 0 \\ 0 & 0 & 0 & 0 & 0 & -t \end{pmatrix} = t^6.$$

Por lo tanto T tiene únicamente un eigenvalor ($\lambda = 0$) y $\mathsf{K}_\lambda = \mathsf{V}$. Sea β cualquier base canónica de Jordan para T. Si r_i es el número de puntos en el renglón i del diagrama de puntos para β entonces $r_1 = 6 - \text{rango}(A)$ $(A) = 6 - 3 = 3$.

Como

$$A^2 = \begin{pmatrix} 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$r_2 = \text{rango}(A) - \text{rango}(A^2) = 3 - 1 = 2$. Entonces dado que $r_1 = 3$, $r_2 = 2$ y como existen seis puntos en el diagrama, se tiene que $r_3 = 1$, por lo que el diagrama de puntos para β es

$$\begin{array}{ccc} \bullet & \bullet & \bullet \\ \bullet & \bullet & \\ \bullet & & \end{array}$$

Concluimos que la forma canónica de Jordan J de T es

$$J = \left(\begin{array}{ccc|cc|c} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Buscaremos ahora una base canónica de Jordan para T. Como la primera columna del diagrama de puntos para β está formada por tres puntos, debemos encontrar un vector x_1 tal que

$$\frac{\partial^2}{\partial x^2}(x_1) \neq 0.$$

Examinando la base $\alpha = \{1, x, y, x^2, y^2, xy\}$ para K_λ, vemos que x^2 es un candidato para x_1. Haciendo $x_1 = x^2$ encontramos que

$$(\mathsf{T} - \lambda \mathsf{I})(x_1) = \mathsf{T}(x_1) = \frac{\partial}{\partial x}(x_1) = 2x \quad \text{y}$$

$$(\mathsf{T} - \lambda \mathsf{I})^2(x_1) = \mathsf{T}^2(x_1) = \frac{\partial^2}{\partial x^2}(x_1) = 2.$$

De la misma manera, dado que la segunda columna del diagrama de puntos para β está formada por dos puntos, debemos encontrar un vector x_2 tal que

$$\frac{\partial}{\partial x}(x_2) \neq 0.$$

Examinando a α con 1, x y x^2 no consideradas (porque están dentro del subespacio generado por el ciclo $\{2, 2x, x^2\}$), vemos que podemos escoger $x_2 = xy$. Así

$$(\mathsf{T} - \lambda\mathsf{I})(x_2) = \mathsf{T}(x_2) = \frac{\partial}{\partial x}(xy) = y.$$

Finalmente, selecciónese $x_3 = y^2$. Así habremos identificado la siguiente base con el diagrama de puntos

$$\begin{array}{lll} \cdot 2 & \cdot y & \cdot y^2 \\ \cdot 2x & \cdot xy & \\ \cdot x^2 & & \end{array}$$

Luego entonces, $\beta = \{2, 2x, x^2, y, xy, y^2\}$ es una base canónica de Jordan para T.

En los tres ejemplos anteriores aprovechamos nuestro ingenio, así como el contexto del problema para encontrar una base canónica de Jordan. El lector será capaz de realizar lo mismo en los ejercicios. Tuvimos éxito en estos casos por el hecho de que las dimensiones de los eigenespacios considerados eran pequeñas. No trataremos, sin embargo, de desarrollar un algoritmo general para calcular una base canónica de Jordan aun cuando se podría formular uno siguiendo los pasos de la demostración de la existencia de tal base (Teorema 6.4).

El siguiente resultado puede ser considerado como un corolario del Teorema 6.7.

Teorema 6.8. *Sean* A *y* B *dos matrices cuadradas del mismo tamaño, cada una con formas canónicas de Jordan calculadas de acuerdo con las convenciones de esta sección. Entonces* A *y* B *son similares si y sólo si tienen la misma forma canónica de Jordan (hasta una permutación de sus eigenvalores).*

DEMOSTRACIÓN. Si A y B tienen la misma forma canónica de Jordan J, entonces A y B son ambas similares a J y por lo tanto son similares entre sí.

Recíprocamente, supóngase que A y B son similares. Entonces A y B deben tener los mismos eigenvalores con las mismas multiplicidades. Sean J_A y J_B respectivamente las formas canónicas de Jordan de A y B para algún orden fijo de sus eigenvalores. Entonces como A es similar a J_A y B es similar a J_B, la hipótesis implica que J_A y J_B son similares. Por lo tanto, de acuerdo con el Ejercicio 19 de la Sección 5.1, existe un operador lineal T en un espacio vectorial dimensionalmente finito V, así como bases β y γ para V tales que $[\mathsf{T}]_\beta = J_A$ y $[\mathsf{T}]_\gamma = J_B$. Entonces J_A y J_B son formas canónicas de Jordan para el mismo operador lineal. Por

lo tanto, como los eigenvalores de A y B están ordenados en la misma forma, el corolario al Teorema 6.7 implica que $J_A = J_B$. ■

Ejemplo 8. Determinaremos cuáles de las siguientes matrices son similares.

$$A = \begin{pmatrix} -3 & 3 & -2 \\ -7 & 6 & -3 \\ 1 & -1 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 & -1 \\ -4 & 4 & -2 \\ -2 & 1 & 1 \end{pmatrix},$$

$$C = \begin{pmatrix} 0 & -1 & -1 \\ -3 & -1 & -2 \\ 7 & 5 & 6 \end{pmatrix}, \quad \text{y} \quad D = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

Obsérvese que A, B y C tienen el mismo polinomio característico $-(t-1)(t-2)^2$, mientras que D tiene a $-t(t-1)(t-2)$ como polinomio característico. Entonces, como matrices similares tienen los mismos polinomios característicos, D no puede ser semejante a A, B o C. Ahora bien, cada una de las matrices A, B y C tiene los mismos eigenvalores $\lambda_1 = 1$ y $\lambda_2 = 2$ con multiplicidades respectivas de 1 y 2. Si J_A, J_B y J_C son respectivamente las formas canónicas de Jordan de A, B y C, con respecto a este orden de sus eigenvalores, entonces

$$J_A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}, \quad J_B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad \text{y} \quad J_C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

Como $J_A = J_C$, A es similar a C, mientras que B no es similar ni a A ni a C.

El lector debería observar que cualquier matriz diagonal es una forma canónica de Jordan, de manera que T es diagonalizable si y sólo si su forma canónica de Jordan es una matriz diagonal. Por lo tanto, si T es un operador diagonalizable en V, cualquier base canónica de Jordan para T es una base para V formada por eigenvectores de T.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.
   (a) La forma canónica de Jordan de una matriz diagonal es la matriz misma.
   (b) Sea T un operador lineal en un espacio vectorial dimensionalmente finito V que tiene una forma canónica de Jordan J. Si β es cualquier base para V, entonces la forma canónica de Jordan para $[T]_\beta$ es J.
   (c) Operadores lineales con el mismo polinomio característico son similares.

(d) Matrices con la misma forma canónica de Jordan son similares.
(e) Toda matriz es similar a su forma canónica de Jordan.
(f) Sea T un operador lineal en un espacio vectorial dimensionalmente finito con polinomio característico $(-1)^n(t-\lambda)^n$. T tiene una forma canónica de Jordan única, sujeto a la convención de que los bloques de Jordan estén ordenados por tamaños decrecientes.
(g) Si un operador tiene una forma canónica de Jordan, entonces existe una base canónica de Jordan única para ese operador.
(h) El diagrama de puntos de cualquier operador lineal que tenga una forma canónica de Jordan es único.

2. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V tal que el polinomio característico de T se descompone en un producto de factores de grado 1. Sean $\lambda_1 = 2$, $\lambda_2 = 4$ y $\lambda_3 = -3$ los distintos eigenvalores de T y supóngase que los diagramas de puntos para la restricción de $T - \lambda_i I$ a K_{λ_i} $(i = 1, 2, 3)$ son los siguientes:

$$\lambda_1 = 2 \qquad \lambda_2 = 4 \qquad \lambda_3 = -3$$

$$\begin{array}{ccc} \bullet & \bullet & \bullet \\ \bullet & \bullet & \\ \bullet & & \end{array} \qquad \begin{array}{cc} \bullet & \bullet \\ \bullet & \\ \bullet & \end{array} \qquad \begin{array}{cc} \bullet & \bullet \end{array}$$

Encontrar la forma canónica de Jordan de T.

3. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V tal que la forma canónica de Jordan de T es

$$\begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

(a) Encontrar el polinomio característico de T.
(b) Encontrar el diagrama de puntos correspondiente a cada eigenvalor de T.
(c) ¿Para cuáles eigenvalores λ_i, si es que hay alguno, se tiene que $E_{\lambda_i} = K_{\lambda_i}$?
(d) Para cada eigenvalor λ_i encontrar el entero positivo más pequeño p_i para el cual $K_{\lambda_i} = N((T - \lambda_i I)^{p_i})$.
(e) Sea U_i la restricción de $T - \lambda_i I$ a K_{λ_i} para cada i. Calcular para cada i:

(i) rango(U_i)
(ii) rango(U_i^2)
(iii) nulidad(U_i)
(iv) nulidad(U_i^2)

4. Para cada una de las siguientes matrices A, encontrar una forma canónica de Jordan J y una matriz Q tal que $J = Q^{-1}AQ$.

(a)
$$A = \begin{pmatrix} -3 & 3 & -2 \\ -7 & 6 & -3 \\ 1 & -1 & 2 \end{pmatrix}$$

(b)
$$A = \begin{pmatrix} 0 & 1 & -1 \\ -4 & 4 & -2 \\ -2 & 1 & 1 \end{pmatrix}$$

(c)
$$A = \begin{pmatrix} 0 & -1 & -1 \\ -3 & -1 & -2 \\ 7 & 5 & 6 \end{pmatrix}$$

(d)
$$A = \begin{pmatrix} 0 & -3 & 1 & 2 \\ -2 & 1 & -1 & 2 \\ -2 & 1 & -1 & 2 \\ -2 & -3 & 1 & 4 \end{pmatrix}$$

Nótese que las matrices de los incisos (a), (b) y (c) son matrices utilizadas en el Ejemplo 8.

5. Sea A una matriz de $n \times n$ cuyo polinomio característico se descompone en un producto de factores de grado 1. Demostrar que A y A^t tienen la misma forma canónica de Jordan y concluir que A y A^t son similares. *Sugerencia:* Para cualquier eigenvalor λ de A y A^t y cualquier entero positivo r, demostrar que rango($(A - \lambda I)^r$) = rango($(A^t - \lambda I)^r$).

6. Sea V el espacio vectorial de las funciones que son combinaciones lineales de e^x, xe^x, x^2e^x y e^{2x}. Defínase T: $V \to V$ mediante $T(f) = f'$ (la derivada de f). Encontrar una forma canónica de Jordan así como una base canónica de Jordan para T.

7. Supóngase que un arreglo de puntos (tales como un diagrama de puntos) tiene k columnas y m renglones y que la columna i del arreglo contiene p_i puntos y el renglón i del arreglo contiene r_i puntos. Si $p_1 \geq p_2 \geq \ldots \geq p_k$, demostrar las siguientes proposiciones:

(a) $m = p_1$ y $k = r_1$.
(b) $p_i = \max\{j\colon\ r_j \geq i\}$ para $1 \leq i \leq k$ y $r_i = \max\{j\colon\ p_j \geq i\}$ para $1 \leq i \leq m$. *Sugerencia:* Utilizar inducción sobre m.
(c) $r_1 \geq r_2 \geq \ldots \geq r_m$.
(d) Conclúyase que el número de puntos en cada columna de un diagrama de puntos queda completamente determinado si se conoce el número de puntos en cada renglón.

Definición. *Un operador lineal* T *en* V *se llama* nilpotente *si* $T^p = T_0$ *para algún entero positivo* p.

8. Demostrar que si T es un operador nilpotente en un espacio vectorial n-dimensional V, entonces el polinomio característico de T es $(-1)^n t^n$. Por lo tanto, el polinomio característico de T se descompone en un producto de factores de grado 1 y T tiene únicamente un eigenvalor (cero) de multiplicidad n. *Sugerencia:* Utilizar inducción sobre n. En el paso general, supóngase que la conclusión es verdadera para todos los espacios vectoriales de dimensión menor que n y síganse los pasos siguientes.

(a) Demostrar que T tiene al menos un eigenvector correspondiente a $\lambda = 0$. Luego $\dim(\mathsf{R}(\mathsf{T})) < \dim(\mathsf{V}) = n$.

(b) Aplicar la hipótesis de inducción al subespacio T-invariante R(T).

(c) Extiéndase una base $\{x_1, x_2, \ldots, x_k\}$ para R(T) hasta una base $\beta = \{x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, x_n\}$ para V.

(d) Demostrar que

$$[\mathsf{T}]_\beta = \begin{pmatrix} B_1 & B_2 \\ O & O' \end{pmatrix},$$

donde O y O' son, respectivamente, matrices nulas de $(n-k) \times k$ y $(n-k) \times (n-k)$.

(e) Dedúzcase que $\det(\mathsf{T} - t\mathsf{I}) = (-1)^n t^n$.

9. Demostrar la recíproca del Ejercicio 8: Si T es un operador lineal en V que tenga a $(-1)^n t^n$, como polinomio característico, entonces T es nilpotente.

10. Dar un ejemplo de un operador lineal T tal que T no sea nilpotente pero que el cero sea el único eigenvalor de T. Caracterizar a todas estas transformaciones.

Definición. *Una matriz* A *de* n $\times$ n *se llama* nilpotente *si* A^p *es igual a la matriz cero de* n $\times$ n *para algún entero positivo* p.

11. Sea $A \in \mathsf{M}_{n\times n}(F)$. Demostrar que $A^p = O$, donde O es la matriz nula de $n \times n$, si y sólo si $(\mathsf{L}_A)^p = \mathsf{T}_0$. Conclúyase que A es nilpotente si y sólo si L_A es nilpotente.

12. Demostrar que cualquier matriz triangular cuadrada con todos los elementos de la diagonal iguales a cero es nilpotente.

13. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V tal que el polinomio característico de T se descomponga en un producto de factores de grado 1. Sean $\lambda_1, \lambda_2, \ldots, \lambda_k$ los distintos eigenvalores de T. Como $\mathsf{V} = \mathsf{K}_{\lambda_1} \oplus \mathsf{K}_{\lambda_2} \oplus \cdots \oplus \mathsf{K}_{\lambda_k}$, podemos definir un mapeo $\mathsf{U}\colon \mathsf{V} \to \mathsf{V}$ de la siguiente manera: Para $x \in \mathsf{V}$, donde $x = x_1 + x_2 + \ldots + x_k$ con $x_i \in \mathsf{K}_{\lambda_i}$, defínase

$$\mathsf{U}(x) = (\mathsf{T} - \lambda_1\mathsf{I})(x_1) + (\mathsf{T} - \lambda_2\mathsf{I})(x_2) + \ldots + (\mathsf{T} - \lambda_k\mathsf{I})(x_k).$$

Demostrar que

(a) U es un operador lineal.
(b) U es nilpotente.
(c) UT = TU.

14. Sean T y U como en el Ejercicio 13. Supóngase que β_i es una base canónica de Jordan de la restricción de T a K_{λ_i}, y sea J_i la forma canónica de Jordan para esta restricción. Entonces $\beta = \beta_1 \cup \beta_2 \cup \ldots \cup \beta_k$ es una base canónica de Jordan para T. Sean $J = [T]_\beta$ y $S = T - U$. Demostrar los siguientes incisos:

(a) $[S]_\beta$ es una matriz diagonal cuyos elementos de la diagonal son idénticos a los elementos de la diagonal de J; esto es, si $D = [S]_\beta$, entonces

$$D_{ij} = \begin{cases} J_{ij} & \text{si } i = j \\ 0 & \text{en cualquier otro caso.} \end{cases}$$

(b) Si $M = [U]_\beta$, entonces

$$M_{ij} = \begin{cases} J_{ij} & \text{si } j = i + 1 \\ 0 & \text{en cualquier otro caso.} \end{cases}$$

(c) $J = D + M$.
(d) $MD = DM$.
(e) Como consecuencia de los incisos (c) y (d) existe una expansión binomial para J. Sea p el entero positivo más pequeño para el cual M^p es igual a la matriz nula. Entonces

$$J^r = D^r + rD^{r-1}M + \frac{r(r-1)}{2!}D^{r-2}M^2 + \ldots + \\ + rDM^{r-1} + M^r \qquad \text{si } r < p,$$

y

$$J^r = D^r + rD^{r-1}M + \frac{r(r-1)}{2!}D^{r-2}M^2 + \ldots + \\ + \frac{r!}{(r-p+1)!(p-1)!}D^{r-p+1}M^{p-1} \qquad \text{si } r \geq p.$$

(f) Si $T = L_A$, entonces existe una matriz Q tal que $A = QJQ^{-1}$.
(g) Para la matriz anterior Q y para cualquier entero positivo r, $A^r = QJ^rQ^{-1}$.

15. Sea T un operador lineal nilpotente en un espacio vectorial dimensionalmente finito V. Recuérdese del Ejercicio 8 que $\lambda = 0$ es el único eigenvalor de T; por lo tanto $V = K_\lambda$. Sea β una base canónica de Jordan para T. Demostrar que para cualquier entero positivo i, si suprimimos de β los vectores correspondientes a los últimos i puntos en cada columna de un

diagrama de puntos para β, el conjunto resultante es una base para $R(T^i)$. (Si una columna del diagrama de puntos contiene menos de i puntos, todos los vectores asociados con esa columna serán eliminados de β.)

16. Encontrar un operador lineal en un espacio vectorial dimensionalmente finito que tenga dos bases canónicas de Jordan distintas.

17. Sea T un operador lineal y sea λ un eigenvalor de T.

(a) Demostrar que $\dim(K_\lambda)$ es la suma de las longitudes de todos los bloques correspondientes a λ en la forma canónica de Jordan de T.

(b) Deducir que $E_\lambda = K_\lambda$ si y sólo si todos los bloques de Jordan correspondientes a λ son de 1×1.

18. (a) Sea J el bloque de Jordan correspondiente al eigenvalor λ de una matriz; entonces

$$J = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix}.$$

Supóngase que J es de $m \times m$ y sea $N = J - \lambda I_m$. Demostrar que N^m es la matriz nula.

(b) Obsérvese, como en el Ejercicio 14, que para cualquier $r \geq m$

$$J^r = \lambda^r I_m + r\lambda^{r-1}N + \frac{r(r-1)}{2!}\lambda^{r-2}N^2 + \ldots + \frac{r(r-1)\ldots(r-m+2)}{(m-1)!}\lambda^{r-m+1}N^{m-1}$$

$$= \begin{pmatrix} \lambda^r & r\lambda^{r-1} & \frac{r(r-1)}{2!}\lambda^{r-2} & \cdots & \frac{r(r-1)\cdots(r-m+2)}{(m-1)!}\lambda^{r-m+1} \\ 0 & \lambda^r & r\lambda^{r-1} & \cdots & \frac{r(r-1)\cdots(r-m+3)}{(m-2)!}\lambda^{r-m+2} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & \lambda^r \end{pmatrix}.$$

Demostrar que existe $\lim_{r \to \infty} J^r$ si y sólo si una de las siguientes condiciones se cumple:

(i) $|\lambda| < 1$.

(ii) $\lambda = 1$ y $m = 1$.

Además, demostrar que $\lim_{r\to\infty} J^r$ es la matriz nula si se cumple (i) y que es la matriz (I) si se cumple (ii).

(c) Demostrar el Teorema 5.16.

19. Para cualquier $A \in \mathsf{M}_{n\times n}(C)$ defínase $\|A\| = \max\{|A_{ij}|: \ 1 \le i, j \le n\}$. Demostrar los siguientes resultados para cualesquiera A, $B \in \mathsf{M}_{n\times n}(C)$ y $c \in C$.

(a) $\|A\| \ge 0$ y $\|A\| = 0$ si y sólo si A es la matriz nula.
(b) $\|cA\| = |c| \cdot \|A\|$.
(c) $\|A + B\| \le \|A\| + \|B\|$.
(d) $\|AB\| \le n\|A\| \cdot \|B\|$.

20. Sean $A \in \mathsf{M}_{n\times n}(R)$ una matriz de transición y $P^{-1}AP = J$ la forma canónica de Jordan de A. Sea $\|\cdot\|$ como se definió en el Ejercicio 19.

(a) Demostrar que para todo entero positivo m, $\|A^m\| \le 1$.
(b) Deducir que $\{\|J^m\|: \ m = 1, 2, \ldots\}$ está acotada.
(c) Utilizando el inciso (b) anterior y el Ejercicio 18(b), demostrar que cada bloque de Jordan correspondiente al eigenvalor $\lambda = 1$ de A es de 1×1.
(d) Utilizando el inciso (c), el Teorema 5.16 y el Ejercicio 18(b), demostrar que $\lim_{m\to\infty} A^m$ existe si y sólo si A tiene la propiedad de que siempre que λ sea un eigenvalor de A con $|\lambda| = 1$, entonces $\lambda = 1$.
(e) Demostrar el Teorema 5.23(a) utilizando al inciso (c) y el Teorema 5.22.

21. (Este ejercicio requiere de conocimientos acerca de series absolutamente convergentes.) Recuérdese de la p. 297 que si $A \in \mathsf{M}_{n\times n}(C)$, entonces e^A se define como $\lim_{m\to\infty} B_m$, donde

$$B_m = I + A + \frac{A^2}{2!} + \ldots + \frac{A^m}{m!}.$$

Utilizar el Ejercicio 19(d) para demostrar que e^A existe para cada $A \in \mathsf{M}_{n\times n}(C)$.

6.3* FORMA CANONICA RACIONAL

A lo largo de los Capítulos 5 y 6 hemos estado utilizando los eigenvalores y los eigenvectores en nuestro análisis sobre operadores lineales en un espacio vectorial dimensionalmente infinito y, como hemos visto, son herramientas útiles siempre y cuando el polinomio característico del operador lineal se descomponga en un producto de factores de grado 1. Exis-

ten, sin embargo, operadores lineales donde éste no es el caso. En efecto, existen operadores lineales sin eigenvalores. Lo que debe hacerse en estos casos es generalizar los conceptos de eigenvalor y de eigenvector con el objeto de obtener teoremas estructurales para reemplazar a los que encontramos en las secciones anteriores.

Dado un operador lineal T en un espacio vectorial dimensionalmente finito V con un polinomio característico $f(t)$, podemos siempre descomponer a $f(t)$ de manera única como un producto de potencias de distintos polinomios mónicos irreducibles multiplicados por $(-1)^n$, donde $n = \dim(\mathsf{V})$. Entonces

$$f(t) = (-1)^n(\phi_1(t))^{n_1}(\phi_2(t))^{n_2} \ldots (\phi_k(t))^{n_k},$$

donde $\phi_i(t)$ es un polinomio mónico irreducible de grado positivo, n_i es un entero positivo $(i = 1, 2, \ldots, k)$ y $\phi_i(t) \neq \phi_j(t)$ para $i \neq j$. Esto se obtiene directamente del teorema de factorización única del Apéndice E. En caso de que $f(t)$ se descomponga en un producto de factores de grado 1, $\phi_i(t) = t - \lambda_i$ para algún eigenvalor $\lambda_i (i = 1, 2, \ldots, k)$. En este caso existe una correspondencia uno-a-uno entre el conjunto de los distintos eigenvalores y el conjunto de los distintos factores mónicos irreducibles del polinomio característico. En el caso general pueden no existir eigenvalores, pero los factores mónicos irreducibles siempre existen. Por esto, es razonable buscar teoremas estructurales basados en los factores mónicos irreducibles del polinomio característico en vez de en los eigenvalores y los eigenvectores.

En esta sección consideraremos algunos teoremas estructurales que pertenezcan a esta situación más general. Para cualquier operador lineal T en un espacio vectorial dimensionalmente finito V, veremos que V puede descomponerse como una suma directa de subespacios T-cíclicos. Además, imponiendo ciertos requerimientos adicionales que relacionen a los subespacios T-cíclicos en la suma con los factores mónicos irreducibles del polinomio característico de T, obtendremos un teorema de unicidad que involucra a algunas propiedades de estos subespacios T-cíclicos. Consecuentemente, será posible escoger una base β para V con el objeto de obtener una matriz $[\mathsf{T}]_\beta$ que sea única para T de la misma manera en que la forma canónica de Jordan de un operador es única para ese operador. Esta matriz será denominada "forma canónica racional" de T. Esta forma podrá utilizarse en lugar de la forma canónica de Jordan en caso de que el polinomio característico de T no se descomponga en un producto de factores de grado 1.

En este punto, sería de mucha ayuda para el lector repasar las definiciones y las técnicas utilizadas en las Secciones 5.4 y 5.5. En particular, el lector debería estudiar los subespacios cíclicos y las matrices compañeras y observar la relación existente entre ellos. Esta relación es tan importante para nuestro desarrollo que deseamos recalcarla en este momento: Dado un operador lineal T en un espacio vectorial dimensionalmente finito V

y un vector no nulo $x \in V$, supóngase que el subespacio T-cíclico de V generado por x, C_x tiene dimensión $d > 0$. Se tiene de la Sección 5.4 que $\beta = \{x, T(x), \ldots, T^{d-1}(x)\}$ es una base ordenada para C_x. Luego $T^d(x)$ es una combinación de β, digamos

$$T^d(x) = -a_0x - a_1T(x) - \ldots - a_{d-1}T^{d-1}(x)$$

para escalares únicos $-a_0, -a_1, \ldots, -a_{d-1}$. (Hemos utilizado $-a_i$ en vez de a_i por conveniencia en la notación.) Como C_x es un subespacio T-invariante de V, podemos considerar la restricción de T a C_x, T_{C_x}. Como ya hemos visto,

$$[T_{C_x}]_\beta = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{d-1} \end{pmatrix}.$$

Esta matriz tiene el polinomio característico

$$f(t) = (-1)^d(a_0 + a_1t + \ldots + a_{d-1}t^{d-1} + t^d)$$

y se denomina *matriz compañera* de $f(t)$. Como consecuencia del Teorema 5.32, el polinomio

$$p(t) = a_0 + a_1t + \ldots + a_{d-1}t^{d-1} + t^d$$

es el polinomio mínimo de T_{C_x}. Además, de acuerdo con el Ejercicio 14 de la Sección 5.6, el polinomio $p(t)$ es el T-aniquilador de x. (A propósito, es esencial para el lector realizar los Ejercicios 13 y 14 de la Sección 5.6 porque serán necesarios para establecer algunos de los resultados de esta sección.)

Considérese de nuevo el operador lineal anterior T. Supóngase que V se descompone en una suma directa de subespacios T-cíclicos

$$V = C_{x_1} \oplus C_{x_2} \oplus \cdots \oplus C_{x_k}$$

para algunos vectores no nulos $x_1, x_2, \ldots, x_k$ en V, donde $\dim(C_{x_i}) = d_i$ para cada i. Si $\beta_i = \{x_i, T(x_i), \ldots, T^{d_i-1}(x_i)\}$ y $\beta = \beta_1 \cup \beta_2 \cup \ldots \cup \beta_k$, entonces β es una base para V y

$$[T]_\beta = [T_{C_{x_1}}]_{\beta_1} \oplus [T_{C_{x_2}}]_{\beta_2} \oplus \cdots \oplus [T_{C_{x_k}}]_{\beta_k}$$

en virtud del Teorema 5.26. Nótese que $[T]_\beta$ es una suma directa de matrices compañeras. Podemos resumir lo anterior de la manera siguiente.

Teorema 6.9. *Sea T un operador lineal en un espacio vectorial dimensionalmente finito V. Si V se puede descomponer como una suma directa de subes-*

pacios T-*cíclicos, entonces existe una base* β *para* V *tal que* $[T]_\beta$ *es una suma directa de matrices compañeras.*

Como enunciamos anteriormente, el objeto de esta sección es demostrar que siempre se puede descomponer a V en una suma directa de subespacios T-cíclicos. Además, se demostrará que siempre es posible escoger cada subespacio T-cíclico C_x tal que el aniquilador de x sea de la forma $(\phi(t))^m$, donde $\phi(t)$ es un factor mónico irreducible del polinomio característico de T y m es un entero positivo. La ventaja de esta descomposición es que la matriz correspondiente a una base seleccionada tal como en el Teorema 6.9 es esencialmente única (sujeta a ciertas convenciones que implican el orden de los subespacios cíclicos).

La exposición anterior nos conduce a la definición siguiente.

Definición. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V *y sea* β *una base ordenada para* V. *La matriz* $[T]_\beta$ *se llamará* forma canónica racional *de* T *si*

$$[T]_\beta = C_1 \oplus C_2 \oplus \cdots \oplus C_k,$$

donde cada C_i *es la matriz compañera de un polinomio de la forma* $(-1)^{md}(\phi(t))^m$, $\phi(t)$ *es un factor mónico irreducible del polinomio característico de* T, d *es el grado de* $\phi(t)$ *y* m *es un entero positivo.*

El siguiente resultado es simplemente otra forma de enunciar el Teorema 6.9 con la terminología anterior.

Corolario. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V. *Si* V *puede descomponerse como una suma directa de subespacios* T-*cíclicos*

$$V = C_{x_1} \oplus C_{x_2} \oplus \cdots \oplus C_{x_k}$$

tales que para cada i, x_i *tiene un aniquilador* $(\phi_i(t))^{m_i}$, *donde* $\phi_i(t)$ *es un factor mónico irreducible del polinomio característico de* T *y* m_i *es un entero positivo, entonces* T *tiene una forma canónica racional.*

En el Teorema 6.5 vimos que si el polinomio característico de un operador lineal T en un espacio vectorial dimensionalmente finito V se descompone en un producto de factores de grado 1, entonces

$$V = K_{\lambda_1} \oplus K_{\lambda_2} \oplus \cdots \oplus K_{\lambda_k},$$

donde $\lambda_1, \lambda_2, \ldots, \lambda_k$ son los distintos eigenvalores de T. Nuestro siguiente resultado será semejante al del Teorema 6.5 en el caso de que el polinomio característico de T no se descomponga en factores de grado 1. Primero, sin embargo, introduciremos el análogo de los eigenespacios generalizados de T.

Definición. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V *con polinomio característico*

$$f(t) = (-1)^n(\phi_1(t))^{n_1}(\phi_2(t))^{n_2} \ldots (\phi_r(t))^{n_r},$$

donde $\phi_1(t), \phi_2(t), \ldots, \phi_r(t)$ *son distintos polinomios mónicos irreducibles y* $n, n_1, n_2, \ldots, n_r$ *son enteros positivos. Para cada* $i = 1, 2, \ldots, r$, *defínase*

$$K_{\phi_i} = N((\phi_i(T))^{n_i}).$$

Obsérvese que para cualquier polinomio $g(t)$, T conmuta con $g(T)$. Por lo tanto, cada K_{ϕ_i} es T-invariante. Además, si $p(t)$ es el polinomio mínimo de T, entonces de acuerdo con el Teorema 5.29

$$p(t) = (\phi_1(t))^{m_1}(\phi_2(t))^{m_2} \ldots (\phi_r(t))^{m_r}$$

para algunos enteros m_i tales que $0 \leq m_i \leq n_i$ para $1 \leq i \leq r$. Veremos más adelante que, de hecho, $m_i \geq 1$ para cada i.

Teorema 6.10. *(Teorema de la descomposición primaria.) Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V *con un polinomio mínimo* $p(t) = (\phi_1(t))^{m_1} \ldots (\phi_r(t))^{m_r}$, *donde* $\phi_1(t), \ldots, \phi_r(t)$ *son los distintos factores mónicos irreducibles de* $p(t)$. *Entonces*

(a) $V = K_{\phi_1} \oplus \ldots \oplus K_{\phi_r}$.

(b) *Para* $1 \leq i \leq r$ *el polinomio mínimo de la restricción de* T *a* K_{ϕ_i} *es* $(\phi_i(t))^{m_i}$.

DEMOSTRACIÓN. La demostración la haremos por inducción sobre r. Si $r = 1$, la conclusión es inmediata. Así, supóngase que el teorema ha sido demostrado para operadores con polinomios mínimos que tienen $r - 1$ distintos factores mónicos irreducibles para algún entero $r > 1$.

Sean $g(t) = (\phi_1(t))^{m_1}$ y $h(t) = (\phi_2(t))^{m_2} \ldots (\phi_r(t))^{m_r}$. Entonces $g(t)$ y $h(t)$ son primos relativos. Demostraremos que $V = W_1 \oplus W_2$, donde $W_1 = N(g(T))$ y $W_2 = N(h(T))$. Como $g(t)$ y $h(t)$ son primos relativos existen polinomios $q(t)$ y $r(t)$ tales que $q(t)g(t) + r(t)h(t) = 1$ donde 1 es el polinomio constante. (Véase Apéndice E.) Al sustituir T en la ecuación se tiene

$$q(T)h(T) + r(T)g(T) = I. \qquad (4)$$

Entonces $v = q(T)h(T)(v) + r(T)g(T)(v)$ para cada $v \in V$. Pero

$$g(T)q(T)h(T)(v) = q(T)h(T)g(T)(v) = q(T)p(T)(v) = q(T)T_0(v) = 0;$$

esto es, $q(T)h(T)(v) \in W_1$. Del mismo modo, $r(T)g(T)(v) \in W_2$. Luego $V = W_1 + W_2$. Finalmente, si $w \in W_1 \cap W_2$, entonces por la ecuación (4)

$$w = I(w) = q(T)h(T)(w) + r(T)g(T)(w) = 0 + 0 = 0;$$

y así $V = W_1 \oplus W_2$.

Obsérvese que como T conmuta con $g(T)$ y $h(T)$, W_1 y W_2 son T-invariantes. Sean $p_1(t)$ y $p_2(t)$ los polinomios mínimos de T_1 y T_2, las restricciones de T a W_1 y W_2, respectivamente. Ahora demostraremos que $p_1(t) = g(t)$ y $p_2(t) = h(t)$. Las definiciones de W_1 y W_2 muestran que $g(T_1)$ y $h(T_2)$ son ambos operadores nulos. Entonces

$$p_1(t) \text{ divide a } g(t) \quad \text{y} \quad p_2(t) \text{ divide a } h(t). \tag{5}$$

Entonces $p_1(t)p_2(t)$ divide a $g(t)h(t) = p(t)$. Pero para cualquier $v \in V$, $v = w_1 + w_2$ para algunos $w_1 \in W_1$ y $w_2 \in W_2$. Por lo tanto

$$p_1(T)p_2(T)(v) = p_2(T)p_1(T)(w_1) + p_1(T)p_2(T)(w_2) = 0 + 0 = 0.$$

Luego $p(t)$, el polinomio mínimo de T, también divide a $p_1(t)p_2(t)$. Como $p(t)$, $p_1(t)$ y $p_2(t)$ son todos mónicos, se tiene que $p(t) = p_1(t)p_2(t)$. Finalmente la ecuación $g(t)h(t) = p_1(t)p_2(t)$, la ecuación (5) y el hecho de que los cuatro polinomios sean mónicos implica que $p_1(t) = g(t)$ y $p_2(t) = h(t)$.

Aplicando la hipótesis de inducción a T_2 y W_2 tenemos que

$$W_2 = K'_{\phi_2} \oplus \cdots \oplus K'_{\phi_r}, \text{ donde } K'_{\phi_i} = N((\phi_i(T_2))^{m_i}),$$

y que $(\phi_i(t))^{m_i}$ es el polinomio mínimo de la restricción de T_2 a K_{ϕ_i}. Pero como $N((\phi_i(T))^{m_i}) \subseteq W_2$, se tiene que

$$K_{\phi_i} = N((\phi_i(T))^{m_i}) = N((\phi_i(T_2))^{m_i}) = K'_{\phi_i} \text{ para } i = 2, \ldots, r.$$

Además, la restricción de T a K_{ϕ_i} es la misma que la restricción de T_2 a $K_{\phi_i}(i = 2, \ldots, r)$, y en consecuencia $(\phi_i(t))^{m_i}$ es el polinomio mínimo de la restricción de T a K_{ϕ_i}. Luego entonces, como

$$V = W_1 \oplus W_2 = K_{\phi_1} \oplus K'_{\phi_2} \oplus \cdots \oplus K'_{\phi_r} = K_{\phi_1} \oplus K_{\phi_2} \oplus \cdots \oplus K_{\phi_r},$$

la demostración está completa. ■

Principiaremos ahora el proceso de demostrar que todo operador lineal T en un espacio vectorial dimensionalmente finito V tiene, sujeto a ciertas convenciones, una forma canónica racional única. Principiaremos con el caso especial en que el polinomio característico de T sea de la forma $\pm(\phi(t))^n$, donde $\phi(t)$ es irreducible y n es un entero positivo. En este caso el polinomio mínimo de T es de la forma $(\phi(t))^m$ para algún entero positivo $m \leq n$ y podemos demostrar que V puede descomponerse en una suma directa de subespacios T-cíclicos.

Teorema 6.11. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V *con polinomio mínimo* $p(t) = (\phi(t))^m$, *donde* $\phi(t)$ *es un polinomio mónico irreducible y* m *es un entero positivo. Entonces existen vectores no nulos* $x_1, x_2, \ldots, x_k$ *en* V *y enteros positivos* $n_1, n_2, \ldots, n_k$ *con* $n_i \leq m$ *para cada* i *tales que*

(a) $V = C_{x_1} \oplus C_{x_2} \oplus \cdots \oplus C_{x_k}$, *donde* C_{x_i} *es el subespacio* T-*cíclico generado por* x_i.

(b) $(\phi(t))^{n_i}$ *es el* T-*aniquilador de* x_i *para* $i = 1, 2, \ldots, k$.

DEMOSTRACIÓN. La demostración se hará por inducción sobre la dimensión de V. Si $\dim(V) = 1$, el resultado es trivial. Supóngase entonces que el teorema se cumple para todos los espacios vectoriales de dimensión menor que n, donde $n > 1$ es un entero, y sea $\dim(V) = n$.

Como el polinomio mínimo de T es $p(t) = (\phi(t))^m$, existe un vector no nulo x_i en V tal que $(\phi(T))^{m-1}(x_1) \neq 0$. Y así el T-aniquilador de x_i es $p(t)$. Sea $W = C_{x_1}$ y recuérdese que W es T-invariante. Sea

$$\bar{T}: V/W \to V/W$$

el operador lineal inducido por T en el espacio cociente V/W. (Véase Ejercicio 13 de la Sección 5.4.) Se puede ver fácilmente que para cualquier polinomio $g(t)$ el operador inducido por $g(T)$ en V/W es $g(\bar{T})$. Por lo tanto, si $g(T) = T_0$, entonces $g(\bar{T}) = \bar{T}_0$. Así el polinomio mínimo de $\bar{T}$ divide a $p(t)$ y por lo tanto la hipótesis de inducción se aplica a $\bar{T}$ y a V/W. En consecuencia existen subespacios $\bar{T}$-cíclicos $C_2, \ldots, C_k$ de V/W tales que

$$V/W = C_2 \oplus \ldots \oplus C_k$$

y tales que para $2 \leq i \leq k$ el $\bar{T}$-aniquilador del generador de C_i es $(\phi(t))^{n_i}$ para algún entero positivo $n_i \leq m$.

Demostraremos que para $2 \leq i \leq k$ existe un vector x_i en el generador de C_i tal que el T-aniquilador de x_i es $(\phi(t))^{n_i}$. Sea y un elemento del generador de C_i; entonces $(\phi(T))^n(y) \in W = C_{x_1}$. Luego entonces existe un polinomio $h(t)$ tal que

$$(\phi(T))^{n_i}(y) = h(T)(x_1). \tag{6}$$

Por el hecho de que $(\phi(t))^m$ es el polinomio mínimo de T, se tiene de la ecuación (6) que

$$0 = (\phi(T))^m(y) = (\phi(T))^{m-n_i}h(T)(x_1).$$

Ahora bien, $(\phi(t))^m$ es el T-aniquilador de x_1. Por lo tanto $(\phi(t))^m$ divide a $(\phi(t))^{m-n_i}h(t)$ y en consecuencia $(\phi(t))^{n_i}$ divide a $h(t)$. Entonces $(\phi(t))^{n_i}q(t) = h(t)$ para algún polinomio $q(t)$. Defínase a $x_i = y - q(T)(x_1)$. Entonces $y - x_i = q(T)(x_1) \in C_{x_1} = W$, y así tenemos que x_i se encuentra en el generador de C_i. Se tiene de aquí que el T-aniquilador del generador de C_i divide al T-aniquilador de x_i. Pero también, por la ecuación (6),

$$\begin{aligned}(\phi(T))^{n_i}(x_i) &= (\phi(T))^{n_i}(y - q(T)(x_1)) \\ &= (\phi(T))^{n_i}(y) - h(T)(x_1) = 0.\end{aligned}$$

Luego el T-aniquilador de x_i es igual a $(\phi(t))^{n_i}$.

Si el grado de $\phi(t)$ es d, entonces $(\phi(t))^{n_i}$ tiene grado dn_i. Luego, como $(\phi(t))^{n_i}$ es tanto el T-aniquilador de x_i como el $\bar{T}$-aniquilador del generador de C_i, el Teorema 5.27 y el Ejercicio 14 de la Sección 5.6 muestran que

$$\beta_i = \{x_i, T(x_i), \ldots, T^{dn_i-1}(x_i)\}$$

y

$$\gamma_i = \{x_i + W, \bar{T}(x_i + W), \ldots, \bar{T}^{dn_i-1}(x_i + W)\}$$

son, respectivamente, bases para C_{x_i} y C_i. Pero como $V/W = C_2 \oplus \ldots \oplus C_k$, $\gamma_2 \cup \ldots \cup \gamma_k$ es una base para V/W. Se tiene entonces que $\beta_1 \cup \beta_2 \cup \ldots \cup \beta_k$ es una base para V. Por lo tanto $V = C_{x_1} \oplus C_{x_2} \oplus \oplus \ldots \oplus C_{x_k}$. ■

El resultado siguiente se obtiene de manera inmediata a partir del teorema anterior y del corolario al Teorema 6.9.

Corolario 1. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito. Si el polinomio característico o mínimo de* T *es de la forma* $\pm(\phi(t))^m$ *para algún polinomio mónico irreducible* $\phi(t)$ *y para algún entero positivo* m, *entonces* T *tiene una forma canónica racional.*

Corolario 2. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V *con polinomio característico*

$$f(t) = (-1)^n(\phi_1(t))^{n_1}(\phi_2(t))^{n_2} \ldots (\phi_k(t))^{n_k},$$

donde $\phi_1(t), \phi_2(t), \ldots, \phi_k(t)$ *son los distintos factores mónicos irreducibles de* f(t). *Entonces, para cada* i, $(\phi_i(t))^{n_i}$ *es el polinomio característico de* T_i, *la restricción de* T *a* K_{ϕ_i}. *Por lo tanto, para cada* i, K_{ϕ_i} *es no nulo y* $\phi_i(t)$ *es un factor del polinomio mínimo de* T.

DEMOSTRACIÓN. Renumerando a $\phi_1(t), \phi_2(t), \ldots, \phi_k(t)$ si fuera necesario, podemos suponer que el polinomio mínimo de T es

$$p(t) = (\phi_1(t))^{m_1}(\phi_2(t))^{m_2} \ldots (\phi_r(t))^{m_r},$$

donde $r \leq k$ y $1 \leq m_i \leq n_i$ para cada $i = 1, 2, \ldots, r$. Sea $f_i(t)$ el polinomio característico de T_i. Como de acuerdo con el teorema de descomposición primaria $V = K_{\phi_1} \oplus K_{\phi_2} \oplus \cdots \oplus K_{\phi_r}$, entonces por el Teorema 5.25 se tiene que $f(t) = f_1(t)f_2(t) \ldots f_r(t)$.

Considérese cualquier i, $1 \leq i \leq r$, y sea d el grado de $\phi_i(t)$. Como el polinomio mínimo de T_i es $(\phi_i(t))^{m_i}$ en virtud del teorema de descomposición primaria, podemos concluir a partir del Corolario 1 del Teorema 6.11 que existe una base β para K_{ϕ_i} tal que

$$[T_i]_\beta = C_1 \oplus C_2 \oplus \cdots \oplus C_s,$$

donde C_i es la matriz compañera de $(-1)^{q_i d}(\phi_i(t))^{q_i}$ para algunos enteros positivos $q_1, q_2, \ldots, q_s$. Luego, si $d_i = q_1 + q_2 + \ldots + q_s$, tenemos que

$$f_i(t) = \det(C_1 - tI) \cdot \det(C_2 - tI) \cdot \cdots \cdot \det(C_s - tI) = (-1)^{dd_i}(\phi_i(t))^{d_i}.$$

Entonces

$$f(t) = f_1(t)f_2(t) \cdots f_r(t) = \epsilon(\phi_1(t))^{d_1}(\phi_2(t))^{d_2} \cdots (\phi_r(t))^{d_r},$$

donde $\epsilon = \pm 1$. Por lo tanto, el teorema de factorización única implica que $r = k$ y $d_i = n_i$ para toda i. En particular, $f_i(t) = \pm(\phi_i(t))^{n_i}$, $K_{\phi_i} \neq \{0\}$ y $\phi_i(t)$ es un factor del polinomio mínimo de T para $i = 1, 2, \ldots, k$. ■

Continuando con el caso especial en que el polinomio mínimo de T tiene la forma $(\phi(t))^m$ para algunos polinomios mónicos irreducibles $\phi(t)$ de grado d, formularemos ahora un teorema de unicidad para la forma canónica racional de T. Con el objeto de formular este resultado adoptaremos de aquí en adelante la convención de que los vectores $x_1, x_2, \ldots, x_k$ del Teorema 6.11 tendrán siempre sus índices de tal modo que $n_1 \geq n_2 \geq \geq \ldots \geq n_k$. Sujetos a esta convención, demostraremos que los enteros $n_1, n_2, \ldots, n_k$ son únicos. De hecho, proporcionaremos un método para calcular estos enteros. En este momento el lector deberá observar que la unicidad de los enteros $n_1, n_2, \ldots, n_k$ implicará la unicidad de la forma canónica racional de T. De hecho, se tiene que la forma canónica racional de T es

$$C_1 \oplus C_2 \oplus \cdots \oplus C_k,$$

donde C_i es la matriz compañera de $(-1)^{dn_i}(\phi(t))^{n_i}$.

Para ayudarnos en el cálculo de los enteros $n_1, n_2, \ldots, n_k$ en el Teorema 6.11 (y por lo tanto, estableciendo su unicidad), introduciremos un nuevo diagrama de puntos correspondiente a la descomposición de V como una suma directa de subespacios cíclicos. A diferencia de los diagramas de puntos de la Sección 6.2, los diagramas que ahora consideraremos no corresponden a las bases para V. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V que tiene a $(\phi(t))^m$ como polinomio mínimo para algún polinomio mónico irreducible $\phi(t)$ y algún entero positivo m. Supóngase también, al igual que en el Teorema 6.11, que $V = C_{x_1} \oplus \ldots \oplus C_{x_k}$ para algunos vectores no nulos $x_1, x_2, \ldots, x_k$ en V y que, para cada i, x_i tiene al aniquilador $(\phi(t))^{n_i}$ para algún entero positivo n_i. Consideremos que los índices de las x_i son tales que $n_1 \geq n_2 \geq \geq \ldots \geq n_k$. El *diagrama de puntos* asociado con la descomposición anterior se define como el arreglo de puntos que consta de k columnas con n_i puntos en la columna i y ordenadas de manera que la columna i principie

en la parte superior y termine después de n_i puntos. Luego, si $k = 3$, $n_1 = 4$, $n_2 = 2$ y $n_3 = 2$, el diagrama de puntos se vería como

$$\begin{matrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & & \\ \cdot & & \end{matrix}$$

Si definimos a r_i como el número de puntos en el renglón i del diagrama de puntos, vemos que los números r_i quedan determinados por la fórmula dada en el Ejercicio 7 de la Sección 6.2. Además, el conocimiento de los números r_i para toda i nos permite calcular los enteros $n_1, n_2, \ldots, n_k$.

El siguiente teorema nos dice que los r_i pueden expresarse en términos de los rangos de algunos operadores, de donde se deduce que los n_i son únicos y por ello el Teorema proporciona un algoritmo para calcularlas.

Teorema 6.12. *Sean* T *y* r_i *como anteriormente. Entonces*

$$r_1 = \frac{1}{d}[dim(V) - rango(\phi(T))]$$

y

$$r_i = \frac{1}{d}[rango((\phi(T))^{i-1}) - rango((\phi(T))^i)] \qquad para\ i > 1,$$

donde d *es el grado de* $\phi(t)$.

DEMOSTRACIÓN. A continuación se da un bosquejo de la demostración; el lector deberá justificar cada paso.

Podemos establecer ambas situaciones simultáneamente adoptando la convención de que

$$(\phi(T))^i = I \qquad \text{si } i = 0.$$

Entonces para cualquier $i \geq 0$

$$R((\phi(T))^i) = C_{(\phi(T))^i(x_1)} \oplus C_{(\phi(T))^i(x_2)} \oplus \cdots \oplus C_{(\phi(T))^i(x_k)},$$

y por lo tanto (aplicando el Ejercicio 14 de la Sección 5.6 a $\phi^{n_i - i}(t)$)

$$\dim(R((\phi(T))^i)) = \sum_{n_i \geq i} d(n_i - i).$$

Así para $i \geq 1$, por el Ejercicio 7 de la Sección 6.2,

$$\begin{aligned} rango((\phi(T))^{i-1}) - rango((\phi(T))^i) &= d[\sum_{n_j \geq i-1} (n_j - (i-1)) - \sum_{n_j \geq i} (n_j - i)] \\ &= d\sum_{n_j \geq i} [(n_j - (i-1)) - (n_j - i)] \\ &= d\sum_{n_j \geq i} 1 = d(\max\{j:\ n_j \geq i\}) = dr_i. \quad \blacksquare \end{aligned}$$

Corolario 1. *Los enteros* $n_1, n_2, \ldots, n_k$ *del Teorema 6.11 son únicos. Esto es, con la notación del Teorema 6.11, si existen vectores no nulos* $x'_1, x'_2, \ldots, x'_r$ *en* V *y enteros positivos* $n'_1, n'_2, \ldots, n'_r$ *tales que*

$$V = C_{x'_1} \oplus C_{x'_2} \oplus \cdots \oplus C_{x'_r},$$

donde x'_i *tiene como aniquilador a* $(\phi(t))^{n'_i}$ *para* $i = 1, 2, \ldots, r$ *y* $n'_1 \geq n'_2 \geq \ldots \geq n'_r$, *entonces* $k = r, n_1 = n'_1, n_2 = n'_2, \ldots, n_k = n'_r$.

Corolario 2. *Sea* T *como en el Teorema 6.11. Suponiendo que se escoge una base* β *para* V *como en el Teorema 6.9, entonces la forma canónica racional de* T, $[T]_\beta$, *es única. De hecho*

$$[T]_\beta = C_1 \oplus C_2 \oplus \cdots \oplus C_k,$$

donde C_i *es la matriz compañera de* $(-1)^{n_i d}(\phi(t))^{n_i} (i = 1, 2, \ldots, k)$.

Ahora definamos la forma canónica racional de una matriz de la manera natural.

Definición. *La* forma canónica racional *de* $A \in M_{n \times n}(F)$ *se define como la forma canónica racional del operador lineal* L_A: $F^n \longrightarrow F^n$.

Ejemplo 9. Considérese la matriz real A de 4×4 definida mediante

$$A = \begin{pmatrix} 0 & -1 & 5 & -3 \\ 1 & 0 & 0 & -1 \\ 0 & 0 & 3 & -2 \\ 0 & 0 & 5 & -3 \end{pmatrix}.$$

Calcularemos la forma canónica racional de A. El polinomio característico de A es

$$f(t) = \det\begin{pmatrix} -t & -1 & 5 & -3 \\ 1 & -t & 0 & -1 \\ 0 & 0 & 3-t & -2 \\ 0 & 0 & 5 & -3-t \end{pmatrix} = (t^2+1)^2.$$

Así, en nuestra notación anterior, $\phi(t) = t^2 + 1$ y $d = 2$. En el diagrama de puntos para A tenemos

$$r_1 = \tfrac{1}{2}[\dim(R^4) - \text{rango}(\phi(A))] = \tfrac{1}{2}(4 - 0) = 2$$

y

$$r_i = \tfrac{1}{2}[\text{rango}((\phi(A))^{i-1}) - \text{rango}((\phi(A))^i)] = \tfrac{1}{2}(0 - 0) = 0$$

para $i > 1$. Luego, el diagrama de puntos para A es

$\cdot \quad \cdot$

Concluimos que $n_1 = 1$ y $n_2 = 1$. Obsérvese que el número de puntos en el diagrama de puntos es $(1/d)\dim(\mathrm{V})$. (En este caso $\dim(\mathrm{V}) = 4$ y $d = 2$.) Por lo tanto existen vectores x_1 y x_2 en R^1 tales que $\mathrm{R}^4 = \mathrm{C}_{x_1} \oplus \mathrm{C}_{x_2}$ y x_1 y x_2 tienen ambas a $\phi(t) = t^2 + 1$ como aniquilador. Dado que la matriz compañera correspondiente a $t^2 + 1$ es

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

concluimos que la forma canónica racional de A es

$$\left(\begin{array}{cc|cc} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{array}\right).$$

Ejemplo 10. Sea A la matriz real de 4×4

$$A = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

Nuevamente calcularemos la forma canónica racional de A. Nótese que A es una forma canónica de Jordan. Como veremos más adelante, la forma canónica racional de A difiere de su forma canónica de Jordan. Puede verse fácilmente que el polinomio característico $f(t)$ de A es $f(t) = (t - 2)^4$. Así, $\phi(t) = t - 2$ y $d = 1$. Ahora bien,

$$r_1 = 4 - \mathrm{rango}(\phi(A)) = 4 - 2 = 2,$$

$$r_2 = \mathrm{rango}(\phi(A)) - \mathrm{rango}((\phi(A))^2) = 2 - 1 = 1,$$

y

$$r_3 = \mathrm{rango}((\phi(A))^2) - \mathrm{rango}((\phi(A))^3) = 1 - 0 = 1.$$

Dado que hay $4 = \dim(\mathrm{V})/d$ puntos en el diagrama de puntos, podemos terminar el cálculo con r_3, y el diagrama de puntos para A es

$$\begin{array}{cc} \cdot & \cdot \\ \cdot & \\ \cdot & \end{array}$$

Concluimos que $n_1 = 3$ y $n_2 = 1$. Así tenemos que existen elementos x_1 y x_2 en R^1 tales que $\mathrm{R}^4 = \mathrm{C}_{x_1} \oplus \mathrm{C}_{x_2}$, x_1 tiene como aniquilador a $(t - 2)^3$ y x_2 tiene como aniquilador a $t - 2$. Como la matriz compañera de $(-$ $-1)^3(t - 2)^3$

es

$$C_1 = \begin{pmatrix} 0 & 0 & 8 \\ 1 & 0 & -12 \\ 0 & 1 & 6 \end{pmatrix}$$

y la matriz compañera de $(-1)(t-2)$ es $C_2 = (2)$, la forma canónica racional de A es

$$C = C_1 \oplus C_2 = \left(\begin{array}{ccc|c} 0 & 0 & 8 & 0 \\ 1 & 0 & -12 & 0 \\ 0 & 1 & 6 & 0 \\ \hline 0 & 0 & 0 & 2 \end{array}\right).$$

Ejemplo 11. Para las matrices A y C del Ejemplo 10, encontraremos una matriz Q tal que $Q^{-1}AQ = C$.

Obsérvese que A y C son similares por el corolario del Teorema 2.27. Por lo tanto, únicamente tenemos que encontrar una base ordenada β para R^4 tal que $[\mathsf{L}_A]_\beta = C$ y luego tomar a Q como la matriz cuyas columnas son los miembros de β. Para encontrar tal base β, necesitamos encontrar vectores no nulos x_1 y x_2 en R^4 tales que x_1 tiene a $(t-2)^3$ como aniquilador, x_2 tiene a $(t-2)$ como aniquilador y $\{x_1, \mathsf{L}_A(x_1), \mathsf{L}_A^2(x_1), x_2\}$ es linealmente independiente. Para empezar, encontremos un elemento de R^4 con aniquilador $(t-2)^3$, esto es, un elemento x_1 tal que $(\mathsf{L}_A - 2\mathsf{I})^3(x_1) = 0$ pero $(\mathsf{L}_A - 2\mathsf{I})^2(x_1) \neq 0$. Si consideramos metódicamente a los miembros de la base estándar $\{e_1, e_2, e_3, e_4\}$, vemos que e_3 tiene esta propiedad. Haciendo

$$x_1 = e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix},$$

encontramos que

$$\mathsf{L}_A(x_1) = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \end{pmatrix} \quad \text{y} \quad \mathsf{L}_A^2(x_1) = \begin{pmatrix} 1 \\ 4 \\ 4 \\ 0 \end{pmatrix}.$$

A continuación escogemos un elemento $x_2 \in \mathsf{R}^4$ linealmente independiente de $\{x_1, \mathsf{L}_A(x_1), \mathsf{L}_A^2(x_1)\}$ y con aniquilador $t-2$. Es evidente que e_4 satisface esta condición. Así,

$$\beta = \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \\ 4 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

es una base para R^4 tal que $[\mathsf{L}_A]_\beta = C$. Por lo tanto, si

$$Q = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 4 & 0 \\ 1 & 2 & 4 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

entonces $Q^{-1}AQ = C$.

Consideraremos ahora el caso general de un operador lineal T en un espacio vectorial dimensionalmente finito para el que el polinomio característico contiene más de un factor irreducible. Combinando los Teoremas 6.11 y 5.26 podemos demostrar fácilmente que T tiene una forma canónica racional.

Teorema 6.13. *Sea* T *cualquier operador lineal en un espacio vectorial dimensionalmente finito* V. *Entonces* T *tiene una forma canónica racional.*

DEMOSTRACIÓN. Supóngase que el polinomio característico de T es

$$(-1)^n(\phi_1(t))^{m_1}(\phi_2(t))^{m_2} \ldots (\phi_r(t))^{m_r},$$

donde $m_i \geq 1$ y los $\phi_i(t)$'s son polinomios mónicos irreducibles distintos. Si $r = 1$ el resultado se sigue del Corolario 1 del Teorema 6.11. De lo contrario, para cada $i = 1, 2, \ldots, r$, T_i, la restricción de T a K_{ϕ_i} tiene, de acuerdo con el Corolario 2 del Teorema 6.11, el polinomio característico $\pm(\phi_i(t))^{m_i}$. Por lo tanto, de acuerdo con el Corolario 1 del Teorema 6.11, existe una base β_i tal que $[\mathsf{T}_i]_{\beta_i} = D_i$ es una forma canónica racional para T_i. Haciendo $\beta = \beta_1 \cup \beta_2 \cup \ldots \cup \beta_r$, es evidente por el Teorema 6.10 que β es una base para V, de modo que por el Teorema 5.26

$$[\mathsf{T}]_\beta = D_1 \oplus D_2 \oplus \cdots \oplus D_r.$$

Así tenemos que $[\mathsf{T}]_\beta$ es una forma canónica racional para T. ■

La demostración del teorema anterior implica la selección de una base para V que garantice una forma canónica racional para T. Dentro del contexto de este resultado podemos enunciar lo siguiente: Si $D = [\mathsf{T}]_\beta$ es la forma canónica racional arriba construida, entonces $D = D_1 \oplus D_2 \oplus \oplus \ldots \oplus D_r$, y para cada $i = 1, 2, \ldots, r$ existe una sucesión de enteros $n_{i1} \geq n_{i2} \geq \cdots \geq n_{ik_i} \geq 1$ tales que

$$D_i = C_{i1} \oplus C_{i2} \oplus \cdots \oplus C_{ik_i},$$

donde C_{ij} es la matriz compañera de $(-1)^{n_{ij}d_j}(\phi_i(t))^{n_{ij}}$ y d_j es el grado de $\phi_j(t)$.

El teorema siguiente garantiza la unicidad de la forma canónica racional de un operador, siempre que éste satisfaga la descripción anterior.

Teorema 6.14. *Sea* T *un operador lineal en un espacio vectorial dimensionalmente finito* V *con polinomio característico*

$$f(t) = (-1)^n(\phi_1(t))^{m_1}(\phi_2(t))^{m_2}\ldots(\phi_r(t))^{m_r},$$

donde los $\phi_i(t)$ *son polinomios mónicos irreducibles distintos y* $m_i \geq 1$ *para toda* i. *Supóngase que* D *es una forma canónica racional para* T *tal que* $D = D_1 \oplus D_2 \oplus \cdots \oplus D_r$ *y para cada* $i = 1, 2, \ldots, r$ *existe una sucesión de enteros* $n_{i1} \geq n_{i2} \geq \cdots \geq n_{ik_i} \geq 1$ *tales que* $D_i = C_{i1} \oplus C_{i2} \oplus \oplus \ldots \oplus C_{ik_i}$ *donde* C_{ij} *es la matriz compañera de* $(-1)^{n_{ij}d_i}(\phi_i(t))^{n_{ij}}$ *y* d_j *es el grado de* $\phi_j(t)$. *Entonces* D *es única en el sentido de que si* D′ *es cualquier otra forma canónica racional para* T *que satisfaga la descripción anterior para sucesiones de enteros* $n'_{i1} \geq n'_{i2} \geq \cdots \geq n'_{ik_i} \geq 1$, *entonces* $D = D'$.

DEMOSTRACIÓN. Dado $D = D_1 \oplus D_2 \oplus \cdots \oplus D_r$, sea β una base para V tal que $[T]_\beta = D$. Supóngase para cada $i = 1, 2, \ldots, r$ que D_i es una matriz de $p_i \times p_i$. Sean β_1 el conjunto ordenado consistente de los primeros p_1 miembros de β, β_2 el conjunto ordenado formado por los siguientes p_2 miembros de β, y así sucesivamente. Para cada $i = 1, 2, \ldots, r$ defínase W_i como el subespacio generado por β_i. En virtud del hecho de que D es una suma directa de las D_i, W_i es un subespacio T-invariante de V y $[T_{W_i}]_{\beta_i} = D_i$. Como el polinomio característico de D_i es un producto de polinomios característicos de las matrices compañeras C_{ij}, D_i debe tener un polinomio característico $\pm(\phi_i(t))^{m_i}$. Por lo tanto también lo debe tener T_{W_i}. Así, por el teorema de Cayley-Hamilton, $(\phi_i(T))^{m_i}(x) = 0$ para toda $x \in W_i$, de manera que $W_i \subseteq K_{\phi_i}$ y por lo tanto

$$\dim(W_i) \leq \dim(K_{\phi_i}) \qquad \text{para } i = 1, 2, \ldots, r. \tag{7}$$

Como

$$\beta = \bigcup_{i=1}^{r} \beta_i \quad \text{y} \quad \beta_i \cap \beta_j = \varnothing \qquad \text{para } i \neq j,$$

tenemos que $V = W_1 \oplus W_2 \oplus \cdots \oplus W_r$. Por lo tanto

$$\dim(V) = \sum_{i=1}^{r} \dim(W_i). \tag{8}$$

Pero por el teorema de descomposición primaria tenemos también que

$$\dim(V) = \sum_{i=1}^{r} \dim(K_{\phi_i}). \tag{9}$$

Luego, por las ecuaciones (7), (8) y (9), concluimos que $\dim(W_i) = \dim(K_{\phi_i})$ para toda i. Así $W_i = K_{\phi_i}$ para cada i y por lo tanto β_i es una base para K_{ϕ_i} $(i = 1, 2, \ldots, r)$. De aquí que

$$D_i = [T_i]_{\beta_i}$$

es una forma canónica racional para T_i, la restricción de T a K_{ϕ_i}. Pero, de acuerdo con el Corolario 2 del Teorema 6.12, D_i es única y se tiene la unicidad de $D = D_1 \oplus D_2 \oplus \cdots \oplus D_r$. ■

Ejemplo 12. Encontraremos la forma canónica racional de la matriz real

$$A = \begin{pmatrix} 0 & 2 & 0 & -6 & 2 \\ 1 & -2 & 0 & 0 & 2 \\ 1 & 0 & 1 & -3 & 2 \\ 1 & -2 & 1 & -1 & 2 \\ 1 & -4 & 3 & -3 & 4 \end{pmatrix}.$$

Si $f(t)$ es el polinomio característico de A, entonces puede demostrarse que $f(t) = -(t^2 + 2)^2(t - 2)$.

Así, $\phi_1(t) = t^2 + 2$ y $\phi_2(t) = t - 2$ son los distintos factores mónicos irreducibles de $f(t)$. Sean $T = L_A$ y T_i la restricción de T a K_{ϕ_i}. Entonces los polinomios característicos respectivos de T_1 y T_2 son, por el Corolario 2 del Teorema 6.11 $(t^2 + 2)^2$ y $-(t - 2)$, respectivamente. Por tanto $\dim(K_{\phi_1}) = 4$ y $\dim(K_{\phi_2}) = 1$. Como la forma canónica racional de T es la suma directa de las formas canónicas racionales de T_1 y T_2, es necesario calcular cada una de éstas.

Para encontrar la forma canónica racional de T_1, debemos aplicar el Teorema 6.12 a T_1. Pero de acuerdo con el Ejercicio 13 podemos, en vez de ello, aplicar el Teorema 6.12 directamente a T: $R^5 \to R^5$. Primero, sin embargo, obsérvese que el número de puntos en el diagrama de puntos para T_1 es $(1/d)\dim(K_{\phi_1}) = \frac{1}{2}(4) = 2$, donde d es el grado de $\phi_1(t)$. Denotando por r_1 al número de puntos en el primer renglón del diagrama de puntos para T_1, el Ejercicio 13 nos muestra que

$$\begin{aligned} r_1 &= \tfrac{1}{2}[\dim(R^5) - \text{rango}(\phi_1(T))] \\ &= \tfrac{1}{2}[5 - \text{rango}(A^2 + 2I)] \\ &= \tfrac{1}{2}\left[5 - \text{rango}\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 6 & -12 & 6 \\ 0 & 0 & 6 & -12 & 6 \\ 0 & 0 & 6 & -12 & 6 \\ 0 & 0 & 12 & -24 & 12 \end{pmatrix}\right] = \tfrac{1}{2}(5 - 1) = 2. \end{aligned}$$

Tenemos entonces que el primer renglón contiene todos los puntos del diagrama de puntos para T_1; esto es, el diagrama de puntos para T_1 es

• •

Concluimos que $n_1 = n_2 = 1$. Así, si D_1 es la forma canónica racional de T_1, entonces

$$D_1 = C_{11} \oplus C_{12},$$

donde

$$C_{11} = C_{12} = \begin{pmatrix} 0 & -2 \\ 1 & 0 \end{pmatrix}.$$

Por lo tanto

$$D_1 = \left(\begin{array}{cc|cc} 0 & -2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -2 \\ 0 & 0 & 1 & 0 \end{array}\right).$$

La situación para T_2 es trivial. Como $\dim(\mathsf{K}_{\phi_2}) = 1$, el diagrama de puntos contiene únicamente un punto. Así si D_2 es la forma canónica racional para T_2, entonces $D_2 = (2)$ y la forma canónica racional de A es

$$D = D_1 \oplus D_2 = \left(\begin{array}{cc|cc|c} 0 & -2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 2 \end{array}\right).$$

El lector deberá darse cuenta de que si hubiéramos escrito $f(t) = -$ $-(t-2)(t^2+2)^2$ y hecho a $\phi_1(t) = -(t-2)$ y a $\phi_2(t) = t^2+2$, entonces nuestro cálculo de la forma canónica racional de A hubiera dado

$$D = \left(\begin{array}{c|cc|cc} 2 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & -2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right).$$

Cualquiera de las formas de D es aceptable. Nótese que, excepto por la permutación de $\phi_1(t)$ y $\phi_2(t)$, D es única.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

   (a) La forma canónica racional de un operador lineal es la suma directa de matrices compañeras.

   (b) Si T es un operador lineal en un espacio vectorial dimensionalmente finito V y β es una base para V tal que $[\mathsf{T}]_\beta$ sea la suma directa de matrices compañeras, entonces $[\mathsf{T}]_\beta$ es una forma canónica racional para T.

(c) Existen matrices cuadradas que no tienen forma canónica racional.
(d) Una matriz cuadrada es similar a su forma canónica racional.
(e) La forma canónica de Jordan y la forma canónica racional de cualquier operador lineal son las mismas.
(f) Para cualquier operador lineal T en un espacio vectorial dimensionalmente finito V, cualquier factor irreducible del polinomio característico de T divide al polinomio mínimo de T.
(g) Sea $\phi(t)$ un divisor mónico irreducible del polinomio característico de un operador lineal T. Los puntos del diagrama de puntos utilizado para calcular la forma canónica racional de T_{K_ϕ} corresponden uno-a-uno con los vectores de una base para K_ϕ.

2. Para cada una de las siguientes matrices encontrar la forma canónica racional.
(a) La matriz real

$$A = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{pmatrix}$$

(b) La matriz real

$$A = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$$

(c) La matriz compleja

$$A = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$$

(d) La matriz real

$$A = \begin{pmatrix} 0 & -7 & 14 & -6 \\ 1 & -4 & 6 & -3 \\ 0 & -4 & 9 & -4 \\ 0 & -4 & 11 & -5 \end{pmatrix}$$

(e) La matriz real

$$A = \begin{pmatrix} 0 & -4 & 12 & -7 \\ 1 & -1 & 3 & -3 \\ 0 & -1 & 6 & -4 \\ 0 & -1 & 8 & -5 \end{pmatrix}$$

3. Demostrar que si T es un operador lineal en un espacio vectorial dimensionalmente finito V de polinomio mínimo $(\phi(t))^m$ para algún entero positivo m, entonces $N((\phi(T))^{m-1})$ es un subespacio T-invariante propio de V.

4. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V de polinomio característico $f(t) = (-1)^n(\phi_1(t))^{m_1}(\phi_2(t))^{m_2} \ldots (\phi_r(t))^{m_r}$,

donde los $\phi_i(t)$'s son polinomios mónicos distintos irreducibles, m_i es un entero positivo para cada i y $n = \dim(V)$. Demostrar que para cualquier $i = 1, 2, \ldots, r$, si d_i es el grado de $\phi_i(t)$, entonces $\dim(K_{\phi_i}) = m_i d_i$.

5. Sea T como en el Ejercicio 4. Considérese cualesquiera i y j tales que $i \neq j$. Demostrar que la restricción de $\phi_j(T)$ a K_{ϕ_i} es uno-a-uno y sobreyectiva.

6. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V de polinomio mínimo $(\phi(t))^m$ para algún polinomio mónico irreducible $\phi(t)$ y para algún entero positivo m. Demostrar que la restricción de T a $R(\phi(T))$ tiene como polinomio mínimo a $(\phi(t))^{m-1}$.

7. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V. Demostrar que la forma canónica racional de T es una matriz diagonal si y sólo si T es diagonalizable.

8. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V de polinomio característico $f(t) = (-1)^n\phi_1(t)\phi_2(t)$, donde $\phi_1(t)$ y $\phi_2(t)$ son polinomios mónicos irreducibles distintos y $n = \dim(V)$.

(a) Demostrar que existen elementos x_1 y x_2 en V tales que x_1 tiene como T-aniquilador a $\phi_1(t)$, x_2 tiene como T-aniquilador a $\phi_2(t)$ y

$$V = C_{x_1} \oplus C_{x_2}.$$

(b) Demostrar que existe un elemento x_3 en V con T-aniquilador $\phi_1(t)\phi_2(t)$ para el que $V = C_{x_3}$.

Así, para asegurarnos que la descomposición de V en una suma directa de subespacios cíclicos es única, debemos exigir que los generadores de los subespacios cíclicos en la suma tengan las potencias de los factores mónicos irreducibles del polinomio característico iguales a las de sus T-aniquiladores.

9. En la notación del Teorema 6.11, demostrar que los índices de las x_i son tales que $n_1 \geq n_2 \geq \ldots \geq n_k$, entonces $n_1 = m$.

10. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V. Suponiendo que la notación es la misma del enunciado del Teorema 6.14, demostrar que el polinomio mínimo de T es

$$p(t) = (\phi_1(t))^{n_{11}}(\phi_2(t))^{n_{21}} \ldots (\phi_r(t))^{n_{r1}}.$$

11. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V. Demostrar que para cualquier polinomio irreducible $\phi(t)$, si $\phi(T)$ no es uno-a-uno en V, entonces $\phi(t)$ divide al polinomio característico de T. *Sugerencia:* Utilizar el Ejercicio 14 de la Sección 5.6.

12. Justificar la siguiente observación hecha en el Ejemplo 9: Si T es un operador lineal en un espacio vectorial dimensionalmente finito V de polinomio

mínimo $(\phi(t))^m$, donde $\phi(t)$ es irreducible, mónico y de grado d, entonces el número de puntos del diagrama de puntos para T es $\dim(V)/d$.

13. Justificar la aplicación del Teorema 6.12 en el Ejemplo 12; esto es, demostrar el siguiente resultado: Sea T un operador lineal en un espacio vectorial dimensionalmente finito V de polinomio característico

$$f(t) = (-1)^n(\phi_1(t))^{m_1}(\phi_2(t))^{m_2} \ldots (\phi_r(t))^{m_r},$$

donde los $\phi_i(t)$ son los distintos factores mónicos irreducibles de $f(t)$, m_i es un entero positivo $(i = 1, 2, \ldots, r)$ y $n = \dim(V)$. Entonces para cualquier $i = 1, 2, \ldots, r$

$$\dim(V) - \text{rango}\,(\phi_i(T)) = \dim(K_{\phi_i}) - \text{rango}\,(\phi(T_{K_{\phi_i}})),$$

y para cualquier entero $j > 1$

$$\text{rango}\,((\phi_i(T))^{j-1}) - \text{rango}\,((\phi_i(T))^j) = \text{rango}\,((\phi_i(T_{K_{\phi_i}}))^{j-1}) - \text{rango}\,((\phi_i(T_{K_{\phi_i}}))^j).$$

Luego si r_j es el número de elementos del renglón j del diagrama de puntos para $T_{K_{\phi_1}}$ entonces

$$r_1 = \frac{1}{d}[\dim(V) - \text{rango}(\phi_i(T))],$$

y

$$r_j = \frac{1}{d}[\text{rango}((\phi_i(T))^{j-1}) - \text{rango}((\phi_i(T))^j)] \qquad \text{para } j > 1,$$

en donde d es el grado de $\phi_i(t)$.

INDICE DE DEFINICIONES PARA EL CAPITULO 6

Capítulo 7

Espacios con producto interior

Muchas de las aplicaciones de las matemáticas están involucradas con el concepto de medición y, por lo tanto, con el de magnitud o tamaño relativo de diversas cantidades. Luego, no es sorprendente de que los campos de los números reales y los complejos que contienen una noción intrínseca de distancia jueguen un papel especial. En este capítulo, consideraremos que todos nuestros espacios vectoriales se encuentran sobre el campo F, donde F representa a R o a C.

Introduciremos la idea de distancia o longitud en los espacios vectoriales obteniendo una estructura mucho más rica, la famosa "estructura de espacio con producto interior". Esta estructura adicional proporcionará aplicaciones a la geometría (Sección 7.8), a la física (Sección 7.4), condicionamiento en los sistemas de ecuaciones (Sección 7.6), aplicaciones a los mínimos cuadrados (Sección 7.10) y formas cuadráticas (Sección 7.11).

7.1 PRODUCTOS INTERIORES Y NORMAS

Muchas de las nociones geométricas tales como ángulo, longitud y perpendicularidad en R^2 y R^3 pueden extenderse a espacios vectoriales reales y complejos más generales. Todas estas ideas están relacionadas con el concepto de "producto interior".

Definición. *Sea* V *un espacio vectorial sobre* F. *Un* producto interior *en* V *es una función que asigna a cada par ordenado de vectores* x *y* y *en* V *un escalar en* F, *representado como* (x, y), *tal que para toda* x, y *y* z *en* V *y toda* c *en* F *se tiene que:*

(a) $(x + z, y) = (x, y) + (z, y)$.
(b) $(cx, y) = c(x, y)$.
(c) $\overline{(x, y)} = (y, x)$, *donde la barra indica conjugación compleja.*
(d) $(x, x) > 0$ *si* $x \neq 0$.

Nótese que (c) se reduce a $(x, y) = (y, x)$ si $F = R$. Las condiciones (a) y (b) simplemente requieren que el producto interior sea lineal en la primera componente.

Se puede comprobar fácilmente que si $a_1, \ldots, a_n \in F$ y $y, x_1, x_2, \ldots, x_n \in V$, entonces

$$\left(\sum_{i=1}^{n} a_i x_i,\ y\right) = \sum_{i=1}^{n} a_i (x_i,\ y).$$

Ejemplo 1. *Sea* $V = F^n$. *Para* $x = (a_1, \ldots, a_n)$ y $y = (b_1, \ldots, b_n)$ defínase

$$(x,\ y) = \sum_{i=1}^{n} a_i \overline{b}_i.$$

$(.,\ .)$ satisface las condiciones de la (a) a la (d) y se denomina *producto interior ordinario* en F^n. (En cursos elementales de álgebra lineal, éste se denomina *producto punto.)*

La verificación de (a) hasta (d) es sencilla. Por ejemplo, si $z = (c_1, \ldots, c_n)$, tenemos para (a)

$$\begin{aligned}(x + z,\ y) &= \sum_{i=1}^{n} (a_i + c_i)\overline{b}_i = \sum_{i=1}^{n} a_i\overline{b}_i + \sum_{i=1}^{n} c_i\overline{b}_i \\ &= (x,\ y) + (z,\ y).\end{aligned}$$

Así para $x = (1 + i,\ 4)$ y $y = (2 - 3i,\ 4 + 5i)$ en C^2 tenemos que $(x, y) = (1 + i)(2 + 3i) + 4(4 - 5i) = 15 - 15i$.

Ejemplo 2. Si (x, y) es un producto interior cualquiera en un espacio vectorial V y $r > 0$, podemos definir otro producto interior mediante la regla $(x, y)' = r(x, y)$. Si se tuviera que $r < 0$ entonces (d) no se satisfaría.

Ejemplo 3. Sea $V = C([0, 1])$ el espacio vectorial de funciones continuas de valor real en $[0, 1]$. Para $f, g \in V$, defínase $(f, g) = \int_0^1 f(t)g(t)\,dt$. Como la integral anterior es lineal en f, (a) y (b) son inmediatas y (c) es trivial. Si $f \neq 0$, entonces la gráfica de f^2 está ubicada sobre el eje x en algún subintervalo de $[0, 1]$ (aquí se utiliza la continuidad), y por lo tanto $(f, f) = \int_0^1 [f(t)]^2\,dt > 0$.

Definición. *Sea* A *una matriz de* $m \times n$ *con elementos de* F. *Definimos la* transpuesta conjugada *(o* adjunta) de A *como la matriz* A^* *de* $n \times m$ *tal que* $(A^*)_{ij} = \overline{A_{ji}}$.

Ejemplo 4. Sea

$$A = \begin{pmatrix} i & 1 + 2i \\ 2 & 3 + 4i \end{pmatrix}.$$

Entonces

$$A^* = \begin{pmatrix} -i & -2 \\ 1-2i & 3-4i \end{pmatrix}.$$

La transpuesta conjugada de una matriz jugará un papel muy importante en el resto de este capítulo. Nótese que si A tiene elementos reales, entonces A^* es sencillamente la transpuesta de A.

Ejemplo 5. Sea $V = M_{n\times n}(F)$ y defínase $(A, B) = \text{tr}(B^*A)$ para A, $B \in V$. (Recuérdese que la traza de una matriz A se define también como $\text{tr}(A) = \sum_{i=1}^{n} A_{ii}$.) Verificaremos que los incisos (a) y (d) de la definición de producto interior se satisfacen y dejaremos los incisos (b) y (c) al lector. Para ello, sea A, B, $C \in V$. Entonces (utilizando el Ejercicio 6 de la Sección 1.3) $(A + B, C) = \text{tr}(C^*(A + B)) = \text{tr}(C^*A + C^*B) = \text{tr}(C^*A) + \text{tr}(C^*B) = (A, C) + (B, C)$. También

$$(A, A) = \text{tr}(A^*A) = \sum_{i=1}^{n} (A^*A)_{ii} = \sum_{i=1}^{n}\sum_{k=1}^{n} (A^*)_{ik}A_{ki}$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{n} \overline{A_{ki}}A_{ki} = \sum_{i=1}^{n}\sum_{k=1}^{n} |A_{ki}|^2.$$

Por lo tanto, si $A \neq O$, entonces $A_{ki} \neq 0$ para algunas k e i. Así $(A, A) > 0$.

Un espacio vectorial V sobre F dotado con un producto interior específico se llama *espacio con producto interior*. Si $F = C$, llamamos a V *espacio complejo con producto interior,* mientras que si $F = R$, llamamos a V *espacio real con producto interior.*

Así, los Ejemplos 1, 3 y 5 también proporcionan ejemplos de espacios con producto interior. *Para el resto de este capítulo,* F^n *será el espacio con producto interior con el producto interior dado en el Ejemplo 1.*

El lector deberá tener la precaución de que dos distintos productos interiores en un espacio vectorial dado, arrojan dos distintos espacios del producto interior.

Un espacio del producto interior muy importante que se parece a C([0, 1]) es el espacio H de funciones continuas de valor complejo definidas en el intervalo $[0, 2\pi]$ con el producto interior

$$(f, g) = \frac{1}{2\pi}\int_0^{2\pi} f(t)\overline{g(t)}\,dt.$$

La razón de la constante $\frac{1}{2\pi}$ será evidente posteriormente. Este espacio del producto interior, que surge a menudo dentro del contexto de situaciones de tipo físico, será examinado más detenidamente en secciones posteriores.

En este momento mencionaremos algunas cuestiones sobre la integración de funciones de valor complejo. Primero, el número imaginario i puede ser considerado como una constante bajo el signo de integración. Segundo, toda función de valor complejo f puede escribirse como $f = f_1 + if_2$, donde f_1 y f_2 son funciones de valor real.

Entonces tenemos que

$$\int f = \int f_1 + i \int f_2 \quad \text{y} \quad \overline{\int f} = \int \bar{f}.$$

De estas propiedades, así como de la suposición de continuidad, se tiene que H es un espacio con producto interior.

Algunas propiedades que se derivan de inmediato de la definición de un producto interior están contenidas en el siguiente teorema.

Teorema 7.1. *Sea* V *un espacio con producto interior. Entonces para* x, y, z $\in$ V *y* c $\in$ F

(a) $(x, y + z) = (x, y) + (x, z)$.
(b) $(x, cy) = \bar{c}(x, y)$.
(c) $(x, x) = 0$ *si y sólo si* $x = 0$.
(d) *Si* $(x, y) = (x, z)$ *para toda* $x \in V$, *entonces* $y = z$.

DEMOSTRACIÓN.

(a) $$(x, y + z) = \overline{(y + z, x)} = \overline{(y, x) + (z, x)} = \overline{(y, x)} + \overline{(z, x)} = (x, y) + (x, z).$$

Las demostraciones de (b), (c) y (d) se dejan como ejercicios. ■

El lector deberá observar que los incisos (a) y (b) del Teorema 7.1 muestran que el producto interior es *lineal conjugado* en la segunda componente.

Con el objeto de generalizar la noción de longitud en R^3 para espacios con producto interior cualesquiera, necesitamos observar únicamente que la longitud de $x = (a, b, c) \in R^3$ está dada por $\sqrt{a^2 + b^2 + c^2} = \sqrt{(x, x)}$. Por lo tanto, damos la siguiente definición.

Definición. *Sea* V *un espacio con producto interior. Para* $x \in V$ *definimos la* norma *(o* longitud*) de* x *mediante* $\|x\| = \sqrt{(x, x)}$.

Ejemplo 6. Sea $V = F^n$. Entonces

$$\|(a_1, \ldots, a_n)\| = \left[\sum_{i=1}^{n} |a_i|^2\right]^{1/2}$$

es la definición Euclidiana de longitud. Nótese que si $n = 1$, tenemos que $\|a\| = |a|$.

Como sería de esperarse, las conocidas propiedades de la longitud en R^3 se satisfacen en general, como se demostrará más adelante.

Teorema 7.2. *Sea* V *un espacio con producto interior. Entonces para toda* x, $y \in V$ *y* $c \in F$ *tenemos*

(a) $||cx|| = |c| \cdot ||x||$.
(b) $||x|| = 0$ *si y sólo si* $x = 0$. *En cualquier caso* $||x|| \geq 0$.
(c) *(Desigualdad de Cauchy-Schwarz)* $|(x, y)| \leq ||x|| \cdot ||y||$.
(d) *(Desigualdad del triángulo)* $||x + y|| \leq ||x|| + ||y||$.

DEMOSTRACIÓN. Dejaremos la demostración de (a) y (b) como ejercicios.

(c) Si $y = 0$, entonces el resultado es inmediato. Así, supóngase que $y \neq 0$. Entonces, para cualquier $c \in F$, tenemos que

$$0 \leq ||x - cy||^2 = (x - cy, x - cy) = (x, x - cy) - c(y, x - cy)$$
$$= (x, x) - \bar{c}(x, y) - c(y, x) + c\bar{c}(y, y).$$

Haciendo

$$c = \frac{(x, y)}{(y, y)},$$

la desigualdad anterior será

$$0 \leq (x, x) - \frac{|(x, y)|^2}{(y, y)} = ||x||^2 - \frac{|(x, y)|^2}{||y||^2},$$

de donde se obtiene (c).

(d)
$$||x + y||^2 = (x + y, x + y) = (x, x) + (y, x) + (x, y) + (y, y)$$
$$= ||x||^2 + 2\,\mathrm{Re}(x, y) + ||y||^2$$
$$\leq ||x||^2 + 2|(x, y)| + ||y||^2$$
$$\leq ||x||^2 + 2||x|| \cdot ||y|| + ||y||^2$$
$$= (||x|| + ||y||)^2,$$

donde $\mathrm{Re}(x, y)$ es la parte real del número complejo (x, y). Nótese que utilizamos al inciso (c) para demostrar (d). ■

El caso en que se da la igualdad en (c) y (d) se considera en el Ejercicio 15.

Ejemplo 7. Para $V = F^n$ podemos emplear (c) y (d) en el producto interior ordinario para obtener las siguientes muy conocidas desigualdades

$$\left| \sum_{i=1}^{n} a_i b_i \right| \leq \left[\sum_{i=1}^{n} |a_i|^2 \right]^{1/2} \left[\sum_{i=1}^{n} |b_i|^2 \right]^{1/2}$$

y

$$\left[\sum_{i=1}^{n}|a_i + b_i|^2\right]^{1/2} \leq \left[\sum_{i=1}^{n}|a_i|^2\right]^{1/2} + \left[\sum_{i=1}^{n}|b_i|^2\right]^{1/2}.$$

El lector podrá recordar de cursos anteriores que para $V = R^3$ o R^2 tenemos que $(x, y) = ||x|| \cdot ||y|| \cos\theta$ donde θ es el ángulo $(0 \leq \theta \leq \pi)$ entre x e y. Esta ecuación implica a (c) de un modo inmediato puesto que $|\cos\theta| \leq 1$. Nótese igualmente que x e y son perpendiculares si y sólo si $\cos\theta = 0$, esto es, si y sólo si $(x, y) = 0$.

Estamos en el momento de poder generalizar la noción de perpendicularidad para espacios con producto interior cualesquiera.

Definiciones. *Sea* V *un espacio con producto interior. Un vector* x *en* V *es un vector unitario si* $||x|| = 1$. *Los vectores* x *y* y *son* ortogonales (perpendiculares) *si* $(x, y) = 0$. *Un subconjunto* S *de* V *es* ortogonal *si cualquier par de elementos distintos de* S *es ortogonal. Finalmente un subconjunto* S *de* V *es* ortonormal *si* S *es ortogonal y está formado únicamente de vectores unitarios.*

Nótese que si $S = \{x_1, x_2, \ldots, x_n\}$, entonces S es ortonormal si y sólo si $(x_i, x_j) = \delta_{ij}$, donde δ_{ij} es la delta de Kronecker. Obsérvese también que para cualquier vector no nulo x, $(1/||x||)x$ es un vector unitario.

Ejemplo 8. El conjunto $S = \{(1, 1), (1, -1)\}$ en F^2 es ortogonal pero no ortonormal; sin embargo

$$S = \left\{\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right)\right\}$$

es ortonormal.

Ejemplo 9. Recuérdese a H (ver la pág. 381). Proporcionaremos un ejemplo muy importante de un subconjunto ortonormal de H al cual regresaremos en ejemplos posteriores. Defínase a $S = \{e^{ijx}: j \text{ es un entero}\}$, donde i es el número imaginario $\sqrt{-1}$. Claramente S es un subconjunto de H. (Recuérdese que $e^{ijx} = \cos jx + i \operatorname{sen} jx$.) Utilizando la propiedad de que $\overline{e^{it}} = e^{-it}$ para cualquier número real t, tenemos para $j \neq k$ que

$$(e^{ijx}, e^{ikx}) = \frac{1}{2\pi}\int_0^{2\pi} e^{ijt}\overline{e^{ikt}}\, dt = \frac{1}{2\pi}\int_0^{2\pi} e^{i(j-k)t}\, dt$$

$$= \frac{1}{2\pi i(j-k)} e^{i(j-k)t}\Big|_0^{2\pi} = 0.$$

nbién tenemos que

$$(e^{ijx}, e^{ijx}) = \frac{1}{2\pi}\int_0^{2\pi} e^{i(j-j)t}\,dt = \frac{1}{2\pi}\int_0^{2\pi} dt = 1.$$

otras palabras $(e^{ijx}, e^{ikx}) = \delta_{jk}$.

Si consideramos a los espacios R^2 y R^3, es geométricamente evidente que conjuntos ortogonales de vectores no nulos son linealmente independientes. El teorema siguiente nos dice que esto es cierto en cualquier espacio ı producto interior.

7.3. *Sea V un espacio con producto interior, y sea S un conjunto ortoıal formado por vectores no nulos. Entonces S es linealmente independnte.*

MOSTRACIÓN. Sean $x_1, \ldots, x_n$ elementos distintos en S y supóngase

$$0 = \sum_{i=1}^{n} a_i x_i.$$

tonces para cualquier j, $1 \leq j \leq n$,

$$0 = (0, x_j) = \left(\sum_{i=1}^{n} a_i x_i,\ x_j\right) = \sum_{i=1}^{n} a_i (x_i, x_j) = a_j \|x_j\|^2$$

esto que $(x_i, x_j) = 0$ para $i \neq j$. Como $x_j \neq 0$, tenemos que $a_j = 0$. r lo tanto, S es linealmente independiente. ■

Este teorema nos dice, por ejemplo, que el espacio vectorial H del Ejem- 9 contiene un conjunto independiente infinito y por lo tanto no es un acio vectorial dimensionalmente finito.

OS

r si las siguientes afirmaciones son verdaderas o falsas.

Un producto interior es una función de valor escalar dentro del conjunto de pares ordenados de vectores.

Un espacio con producto interior debe estar sobre el campo de los números reales o complejos.

Un producto interior es lineal en ambas componentes.

Existe exactamente un producto interior en el espacio vectorial R^n.

La desigualdad del triángulo sólo se cumple para espacios con producto interior dimensionalmente finitos.

Todo conjunto ortogonal es linealmente independiente.

Todo conjunto ortonormal es linealmente independiente.

(h) Unicamente las matrices cuadradas tienen una transpuesta conjugada.
(i) Si $(x, y) = 0$ para toda x en un espacio con producto interior, entonces $y = 0$.

2. Sea $V = C^3$ con el producto interior ordinario. Sean $x = (2, 1 + i, i)$ y $y = (2 - i, 2, 1 + 2i)$. Calcular (x, y), $||x||$, $||y||$ y $||x + y||^2$. Luego verificar tanto la desigualdad de Cauchy como la del triángulo.

3. En $C([0, 1])$ sea $f(t) = t$ y $g(t) = e^t$. Calcular (f, g) (tal como se definió en el Ejemplo 3), $||f||$, $||g||$ y $||f + g||$. Luego verificar la desigualdad de Cauchy y la del triángulo.

4. Sea $V = M_{n\times n}(F)$ con $(A, B) = \text{tr}(B^*A)$. Completar la demostración del Ejemplo 5 de que $(.,.)$ es un producto interior. Si $n = 2$ y

$$A = \begin{pmatrix} 1 & 2+i \\ 3 & i \end{pmatrix} \text{ y } B = \begin{pmatrix} 1+i & 0 \\ i & -i \end{pmatrix},$$

calcular $||A||$, $||B||$ y (A, B).

5. Demostrar que $(x, y) = xAy^*$ es un producto interior en C^2, donde

$$A = \begin{pmatrix} 1 & i \\ -i & 2 \end{pmatrix}.$$

Calcular (x, y) para $x = (1 - i, 2 + 3i)$ y $y = (2 + i, 3 - 2i)$.

6. Completar la demostración del Teorema 7.1.

7. Completar la demostración del Teorema 7.2.

8. Dar razones por las cuales cada uno de los siguientes incisos *no* son productos interiores en los espacios vectoriales dados.

(a) $((a, b), (c, d)) = ac - bd$ en R^2
(b) $(A, B) = \text{tr}(A + B)$ en $M_{2\times 2}(R)$
(c) $(f, g) = \int_0^1 f'(t)g(t)\,dt$ en $P(R)$, donde $'$ denota diferenciación.

9. Sea β una base para un espacio con producto interior dimensionalmente finito. Demostrar que si $(x, y) = 0$ para toda $x \in \beta$, entonces $y = 0$.

10.* Sea V un espacio con producto interior, supóngase que x y y son elementos ortogonales de V. Demostrar que $||x + y||^2 = ||x||^2 + ||y||^2$. Deducir el teorema de Pitágoras para R^2.

11. Demostrar la *ley del paralelogramo* en un espacio con producto interior V; esto es, demostrar que

$$||x + y||^2 + ||x - y||^2 = 2||x||^2 + 2||y||^2 \text{ para toda } x, y \in V.$$

¿Qué expresa esta ecuación con respecto a paralelogramos en R^2?

12.* Sea $\{x_1, \ldots, x_k\}$ un conjunto ortogonal en V y sean $a_1, \ldots, a_k \in F$. Demostrar que

$$\left\| \sum_{i=1}^{k} a_i x_i \right\|^2 = \sum_{i=1}^{k} |a_i|^2 \|x_i\|^2.$$

13. Supóngase que $(.,\ .)_1$ y $(.,\ .)_2$ son dos productos interiores en un espacio vectorial V. Demostrar que $(.,\ .) = (.,\ .)_1 + (.,\ .)_2$ es otro producto interior en V.

14. Sean A y B matrices de $n \times$ y sea $c \in F$. Demostrar que $(A + cB)^* = A^* + \bar{c}B^*$.

15. (a) Demostrar que si V es ın espacio con producto interior, entonces $|(x, y)| = \|x\| \cdot \|y\|$ si y ólo si uno de los vectores x o y es múltiplo del otro. *Sugerencia:* Si $y \neq 0$, sea

$$a = \frac{(x, y)}{\|y\|^2}.$$

Entonces $x = ay + z$, donde $(y, z) = 0$. Por suposición

$$|a| = \frac{\|x\|}{\|y\|}$$

Aplicar el Ejercicio 10 a $\|x\|^2 = \|ay + z\|^2$ y obténgase $\|z\| = 0$.

(b) Obtener un resultado semejante para la igualdad $\|x + y\| = \|x\| + \|y\|$ y generalizarla para el caso de n vectores.

16. Sea $V = C([0, 1])$, y defínase

$$(f, g) = \int_0^{1/2} f(t)g(t)dt.$$

¿Es éste un producto interior sobre V?

17. Sea V un espacio con producto interior, y supóngase que $T\colon V \to V$ es lineal y que $\|T(x)\| = \|x\|$ para toda x. Demostrar que T es uno-a-uno.

18. Sea V un espacio vectorial sobre F, donde $F = R$ o C, y sea W un espacio con producto interior sobre F con un producto interior $(.,\ .)$. Si $T\colon V \to W$ es lineal, demostrar que $(x, y)' = (T(x), T(y))$ define un producto interior en V si y sólo si T es uno-a-uno.

19. Sea V un espacio con producto interior; demostrar que

(a) $\|x \pm y\|^2 = \|x\|^2 \pm 2\ \mathrm{Re}(x, y) + \|y\|^2$ para toda $x, y \in V$, donde $\mathrm{Re}(x, y)$ es la parte real del número complejo (x, y).

(b) $\big|\|x\| - \|y\|\big| \leq \|x - y\|$ para toda $x, y \in V$.

20. Sea V un espacio sobre F con producto interior. Verificar las *identidades polares*. Para toda $x, y \in V$

(a) $(x, y) = \frac{1}{4}||x + y||^2 - \frac{1}{4}||x - y||^2$ si $F = R$.

(b) $(x, y) = \frac{1}{4}\sum_{k=1}^{4} i^k||x + i^k y||^2$ si $F = C$.

Donde $i = \sqrt{-1}$

21. Sea A una matriz de $n \times n$. Defínase

$$A_1 = \frac{1}{2}(A + A^*) \quad \text{y} \quad A_2 = \frac{1}{2i}(A - A^*).$$

(a) Demostrar que $A_1^* = A_1$, $A_2^* = A_2$ y $A = A_1 + iA_2$. ¿Sería razonable definir a A_1 y A_2 como las partes real e imaginaria, respectivamente, de la matriz A?

(b) Sea A una matriz de $n \times n$. Demostrar que si $A = B_1 + iB_2$ donde $B_1^* = B_1$ y $B_2^* = B_2$, entonces $B_1 = A_1$ y $B_2 = A_2$.

22. Sea V un espacio vectorial sobre F, donde F es R o C. Sea o no V un espacio con producto interior, podemos aún definir una "norma" $||\cdot||$ como una función de valor real en V que satisface las siguientes condiciones para toda $x, y \in V$ y $a \in F$.

(i) $||x|| \geq 0$ y $||x|| = 0$ si y sólo si $x = 0$.

(ii) $||ax|| = |a| \cdot ||x||$.

(iii) $||x + y|| \leq ||x|| + ||y||$.

Demostrar que las siguientes son normas en los espacios vectoriales dados V.

(a) $V = M_{m\times n}(F)$; $\quad ||A|| = \max_{i,j} |A_{ij}|$ para toda $A \in V$

(b) $V = C([0, 1])$; $\quad ||f|| = \max_{t\in[0,1]} |f(t)|$ para toda $f \in V$

(c) $V = C([0, 1])$; $\quad ||f|| = \int_0^1 |f(t)|\, dt$ para toda $f \in V$

(d) $V = R^2$; $\quad ||(a, b)|| = \max\{|a|, |b|\}$ para toda (a, b) en V

Utilizar al Ejercicio 20 para demostrar que no hay ningún producto interior $(.,.)$ en R^2 tal que $||x||^2 = (x, x)$ para toda $x \in R^2$ si $(.,.)$ se define como el inciso (d).

23. Sea V un espacio con producto interior y defínase para cada par ordenado de vectores el escalar $d(x, y) = ||x - y||$ llamado la *distancia* entre x e y. Demostrar, para toda $x, y, z \in V$, que

(a) $d(x, y) \geq 0$.

(b) $d(x, y) = d(y, x)$.

(c) $d(x, y) \leq d(x, z) + d(z, y)$.
(d) $d(x, x) = 0$.
(e) $d(x, y) \neq 0$ si $x \neq y$.

24. Sea V un espacio vectorial real o complejo (posiblemente dimensionalmente infinito), y sea β una base para V. Para $x, y \in \mathsf{V}$ existen $x_1, \ldots, x_n \in \beta$ tales que

$$x = \sum_{i=1}^{n} a_i x_i \quad \text{y} \quad y = \sum_{i=1}^{n} b_i x_i.$$

Defínase

$$(x, y) = \sum_{i=1}^{n} a_i \bar{b}_i.$$

Demostrar que $(.,.)$ es un producto interior en V. Así pues, todo espacio vectorial real o complejo puede ser considerado como un espacio con producto interior.

Demostrar que si $\mathsf{V} = \mathsf{R}^n$ o C^n y β es la base ordenada estándar, entonces el producto interior definido anteriormente es el producto interior ordinario.

7.2 EL PROCESO DE ORTOGONALIZACION DE GRAM-SCHMIDT Y COMPLEMENTOS ORTOGONALES

En capítulos anteriores vimos el papel especial que las bases ordenadas estándar juegan en R^n. Las propiedades de estas bases se derivan del hecho de que los vectores de la base forman un conjunto ortonormal. Así como las bases son los "tabiques" con los que se construyen los espacios vectoriales, las bases que son también conjuntos ortonormales son los "tabiques" de los espacios con producto interior. Ahora daremos nombre a estas bases.

Definición. *Sea* V *un espacio con producto interior. Un subconjunto* β *de* V *es una* base ortonormal *para* V *si* β *es una base ordenada ortonormal.*

Ejemplo 10. Si $\mathsf{V} = \mathsf{F}^n$, entonces la base ordenada estándar es una base ortonormal para V.

Ejemplo 11.

$$\left\{ \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}} \right), \left(\frac{2}{\sqrt{5}}, \frac{-1}{\sqrt{5}} \right) \right\}$$

es una base ortonormal para R^2.

Por supuesto, aún no hemos demostrado que todo espacio del producto interior dimensionalmente finito posee una base ortonormal. El siguiente teorema nos lleva la mayor parte del camino para la obtención de este resultado. Nos dice cómo construir un conjunto ortogonal a partir de un conjunto linealmente independiente de vectores, de tal modo que ambos conjuntos generan el mismo subespacio.

Antes de enunciar este teorema consideremos un caso sencillo. Supóngase que $\{y_1, y_2\}$ es un subconjunto linealmente independiente de un espacio con producto interior (y, por lo tanto, una base para algún subespacio bidimensional). Nos gustaría construir un conjunto ortogonal a partir de $\{y_1, y_2\}$ que genere al mismo subespacio. La figura 7.1 siguiente sugiere que el conjunto $\{x_1, x_2\}$ donde $x_1 = y_1$ y $x_2 = y_2 - cy_1$ darán resultado si se escoge adecuadamente a c.

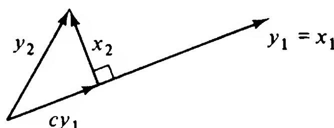


figura 7.1

Para encontrar a c necesitamos únicamente resolver la ecuación siguiente.

$$0 = (x_2, y_1) = (y_2 - cy_1, y_1) = (y_2, y_1) - c(y_1, y_1)$$

Luego

$$c = \frac{(y_2, y_1)}{||y_1||^2}.$$

Y entonces

$$x_2 = y_2 - \frac{(y_2, y_1)}{||y_1||^2} y_1.$$

Este proceso puede extenderse a cualquier subconjunto finito linealmente independiente.

Teorema 7.4. *Sea* V *un espacio con producto interior, y sea* $S = \{y_1, \ldots, y_n\}$ *un subconjunto de* V *linealmente independiente. Defínase* $S' = \{x_1, \ldots, x_n\}$, *donde* $x_1 = y_1$ *y*

$$x_k = y_k - \sum_{j=1}^{k-1} \frac{(y_k, x_j)}{||x_j||^2} x_j \qquad \textit{para } 2 \le k \le n. \tag{1}$$

Entonces S′ *es un conjunto ortogonal de vectores no nulos tales que* $L(S') = L(S)$.

DEMOSTRACIÓN. La demostración se hará por inducción sobre n. Sea $S_n = \{y_1, \ldots, y_n\}$. Si $n = 1$ entonces el teorema se demuestra haciendo

$S'_1 = S_1$; esto es, $x_1 = y_1 \neq 0$. Supóngase luego que el conjunto $S'_k = \{x_1, \ldots, x_k\}$ ha sido construido mediante el uso de la ecuación (1) con las propiedades necesarias. Demostraremos que el conjunto $S'_{k+1} = \{x_1, \ldots, x_k, x_{k+1}\}$ tiene también las propiedades deseadas, donde

$$x_{k+1} = y_{k+1} - \sum_{j=1}^{k} \frac{(y_{k+1}, x_j)}{||x_j||^2} x_j.$$

Si $x_{k+1} = 0$, entonces la ecuación (2) implicaría $y_{k+1} \in L(S'_k) = L(S_k)$, lo que contradice la suposición de que S_{k+1} es linealmente independiente.

Para $1 \leq i \leq k$ tenemos de la ecuación (2) que

$$\begin{aligned}(x_{k+1}, x_i) &= (y_{k+1}, x_i) - \sum_{j=1}^{k} \frac{(y_{k+1}, x_j)}{||x_j||^2}(x_j, x_i)\\ &= (y_{k+1}, x_i) - \frac{(y_{k+1}, x_i)}{||x_i||^2}||x_i||^2 = 0,\end{aligned}$$

puesto que, de acuerdo con la suposición de que S'_k es ortogonal, $(x_j, x_i) = 0$ si $i \neq j$. Por lo tanto S'_{k+1} es ortogonal. Ahora bien, mediante la ecuación (2) tenemos que $L(S'_{k+1}) \subseteq L(S_{k+1})$. Pero de acuerdo con el Teorema 7.3 S'_{k+1} es linealmente independiente; luego $\dim(L(S'_{k+1})) = k + 1 = \dim(L(S_{k+1}))$. Por lo tanto $L(S'_{k+1}) = L(S_{k+1})$. ■

La construcción de $\{x_1, \ldots, x_n\}$ usando la ecuación (1) se llama *proceso de ortogonalización de Gram-Schmidt.*

Ejemplo 12. Sea $V = R^3$ y sean $y_1 = (1, 1, 0)$, $y_2 = (2, 0, 1)$ e $y_3 = (2, 2, 1)$. Entonces $\{y_1, y_2, y_3\}$ es linealmente independiente. Utilizaremos la ecuación (1) anterior para calcular los vectores ortogonales x_1, x_2 y x_3. Tómese $x_1 = (1, 1, 0)$. Entonces $||x_1||^2 = 2$, y así

$$\begin{aligned}x_2 &= y_2 - \frac{(y_2, x_1)}{||x_1||^2} x_1\\ &= (2, 0, 1) - \frac{2}{2}(1, 1, 0)\\ &= (1, -1, 1).\end{aligned}$$

Finalmente,

$$\begin{aligned}x_3 &= y_3 - \frac{(y_3, x_1)}{||x_1||^2} x_1 - \frac{(y_3, x_2)}{||x_2||^2} x_2\\ &= (2, 2, 1) - \frac{4}{2}(1, 1, 0) - \frac{1}{3}(1, -1, 1)\\ &= \left(-\frac{1}{3}, \frac{1}{3}, \frac{2}{3}\right).\end{aligned}$$

Teorema 7.5. *Sea V un espacio dimensionalmente finito con producto interior. Entonces V tiene una base ortonormal β. Además, si $\beta = \{x_1, x_2, \ldots, x_n\}$*

y $x \in V$, *entonces*

$$x = \sum_{i=1}^{n} (x, x_i)x_i.$$

DEMOSTRACIÓN. Sea β_0 una base ordenada para V. Aplicando el Teorema 7.4 para obtener un conjunto ortogonal β' de vectores no nulos con $L(\beta') = L(\beta_0) = V$. Dividiendo cada vector de β' entre su longitud, obtenemos un conjunto ortonormal β que genera a V. De acuerdo con el Teorema 7.3, β es linealmente independiente y, por lo tanto, β es una base ortonormal para V.

Sea $\beta = \{x_1, \ldots, x_n\}$ y sea $x \in V$. Entonces

$$x = \sum_{i=1}^{n} a_i x_i$$

para algunos escalares a_i. Para $1 \leq j \leq n$ tenemos que

$$(x, x_j) = \left(\sum_{i=1}^{n} a_i x_i, x_j\right) = \sum_{i=1}^{n} a_i (x_i, x_j)$$

$$= \sum_{i=1}^{n} a_i \delta_{ij} = a_j. \quad \blacksquare$$

Ejemplo 13. Utilizando el conjunto ortogonal obtenido en el Ejemplo 12, podemos obtener la base ortonormal

$$\left\{\frac{1}{\sqrt{2}}(1, 1, 0), \frac{1}{\sqrt{3}}(1, -1, 1), \frac{1}{\sqrt{6}}(-1, 1, 2)\right\}.$$

Sea $x = (2, 1, 3)$. Calcularemos los "coeficientes" de x como se dan en el Teorema 7.5:

$$a_1 = \frac{1}{\sqrt{2}}(2 + 1) = \frac{3}{\sqrt{2}}, \qquad a_2 = \frac{1}{\sqrt{3}}(2 - 1 + 3) = \frac{4}{\sqrt{3}},$$

y

$$a_3 = \frac{1}{\sqrt{6}}(-2 + 1 + 6) = \frac{5}{\sqrt{6}}$$

Como verificación, tenemos que

$$(2, 1, 3) = \frac{3}{2}(1, 1, 0) + \frac{4}{3}(1, -1, 1) + \frac{5}{6}(-1, 1, 2).$$

Así pues tenemos una manera muy sencilla de calcular los coeficientes de un vector dado cuando se expresa como una combinación lineal de vectores en una base ortonormal.

El mismo teorema proporciona un método sencillo para obtener la representación matricial de un operador lineal.

Corolario. *Sea* V *un espacio dimensionalmente finito con producto interior con una base ortonormal* $\beta = \{x_1, \ldots, x_n\}$. *Sea* T *un operador ortonormal en* V, *y sea* $A = [T]_\beta$. *Entonces* $A_{ij} = (T(x_j), x_i)$.

DEMOSTRACIÓN. Del Teorema 7.5 tenemos que

$$T(x_j) = \sum_{i=1}^{n} (T(x_j), x_i)x_i.$$

Por lo tanto $A_{ij} = (T(x_j), x_i)$. ■

Los escalares (x, x_i) asociados con x han sido estudiados extensivamente para ciertos espacios vectoriales especiales. Aun cuando los vectores $x_1, \ldots, x_n$ se escogieron de una base ortonormal, consideraremos conjuntos β más generales para la definición de los escalares (x, x_i).

Definición. *Sea* β *un subconjunto ortonormal (posiblemente infinito) de un espacio con producto interior* V, *y sea* $x \in V$. *Definimos los* coeficientes de Fourier de x relativos a β *como los escalares* (x, y), *donde* $y \in \beta$.

En el siglo diecinueve el matemático francés Jean Baptiste Fourier estuvo dedicado al estudio de los coeficientes

$$\int_0^{2\pi} f(t) \operatorname{sen} nt\, dt \quad \text{y} \quad \int_0^{2\pi} f(t) \cos nt\, dt,$$

o, más generalmente,

$$c_n = \frac{1}{2\pi} \int_0^{2\pi} f(t)e^{-int}\, dt,$$

de una función f. En el contexto del Ejemplo 9, vemos que $c_n = (f, e^{inx})$; esto es, c_n es el n-ésimo coeficiente de Fourier de una función continua $f \in H$ relativo a S. Estos coeficientes son los coeficientes "clásicos" de Fourier de una función; la literatura concerniente al comportamiento de estos coeficientes es bastante extensa. Aprenderemos más sobre estos coeficientes de Fourier en el resto de este capítulo.

Ejemplo 14. Sean $V = H$ y $f(x) = x$. Calculemos los coeficientes de Fourier de f relativos al conjunto ortonormal S del Ejemplo 9. Utilizando la integración por partes tenemos, para $n \neq 0$,

$$(f, e^{inx}) = \frac{1}{2\pi} \int_0^{2\pi} te^{int}\, dt = \frac{1}{2\pi} \int_0^{2\pi} te^{-int}\, dt = \frac{-1}{in}.$$

Y para $n = 0$

$$(f, 1) = \frac{1}{2\pi} \int_0^{2\pi} t(1)dt = \pi.$$

Ahora bien, de acuerdo con el Ejercicio 14 tenemos que

$$||f||^2 \geq \sum_{n=1}^{k} |(f, e^{inx})|^2$$

para toda k. Luego, empleando el hecho de que $||f||^2 = \frac{4}{3}\pi^2$, tenemos

$$\frac{4}{3}\pi^2 \geq \sum_{n=1}^{k} \left| \frac{-1}{in} \right|^2 = \sum_{n=1}^{k} \frac{1}{n^2}.$$

Como esta desigualdad es cierta para toda k, tenemos mediante el uso adecuado de los límites que

$$\frac{4}{3}\pi^2 \geq \sum_{n=1}^{\infty} \frac{1}{n^2}.$$

Usando otras funciones pueden obtenerse otros resultados semejantes.

Estamos ya preparados para proceder con el concepto de un "complemento ortogonal".

Definición. *Sea* V *un espacio con producto interior y sea* S *un subconjunto de* V. *Definimos a* $S^\perp$ *como el conjunto de todos aquellos vectores de* V *que son ortogonales a todos los vectores de* S; *esto es,* $S^\perp = \{x \in V: (x, y) = 0$ *para toda* $y \in S\}$. *A* $S^\perp$ *se le llama* complemento ortogonal de *S*.

Es fácil demostrar que $S^\perp$ es un subespacio de V para cualquier subconjunto S de V.

Ejemplo 15. El lector deberá verificar que $\{0\}^\perp = V$ y $V^\perp = \{0\}$.

Ejemplo 16. Si $V = R^3$ y $S = \{x\}$, entonces $S^\perp$ es sencillamente el conjunto de todos los vectores que son perpendiculares a x. (Ver el Ejercicio 5.)

El Ejercicio 16 proporciona un ejemplo interesante de un complemento ortogonal en el caso en que V sea dimensionalmente infinito.

El uso de la palabra "complemento" se aclarará con el siguiente teorema.

Teorema 7.6. *Sea* W *un subespacio dimensionalmente finito de un espacio con producto interior* V. *Entonces* $V = W \oplus W^\perp$.

DEMOSTRACIÓN. Por el Teorema 7.5 podemos escoger una base ortonormal $\{x_1, \ldots, x_k\}$ para W. Entonces, para $y \in V$, defínase

$$y_1 = \sum_{i=1}^{k} (y, x_i)x_i \quad \text{y} \quad y_2 = y - y_1.$$

Es evidente que $y = y_1 + y_2$ y $y_1 \in W$. Con el objeto de demostrar que $V = W + W^\perp$, debemos demostrar que $y_2 \in W^\perp$, para lo cual es suficiente demostrar que $(y_2, y_j) = 0$ para $j = 1, \ldots, k$. Ahora bien,

$$(y_2, x_j) = (y - y_1, x_j) = (y, x_j) - (y_1, x_j).$$

Pero

$$(y_1, x_j) = \left(\sum_{i=1}^{k} (y, x_i)x_i, x_j\right) = \sum_{i=1}^{k} (y, x_i)(x_i, x_j) = \sum_{i=1}^{k} (y, x_i)\delta_{ij} = (y, x_j).$$

Por lo tanto $(y_2, y_j) = 0$.

Para completar la demostración debemos demostrar que $W \cap W^\perp = \{0\}$. Pero si $x \in W \cap W^\perp$, entonces $(x, x) = 0$. Por lo tanto $x = 0$. ■

El siguiente resultado es una consecuencia inmediata de la demostración del Teorema 7.6.

Corolario 1. *Bajo la hipótesis del Teorema 7.6, si* $\{x_1, \ldots, x_k\}$ *es una base ortonormal para* W *y si* $y \in V$, *entonces*

$$y = \sum_{i=1}^{k} (y, x_i)x_i + z,$$

donde $z \in W^\perp$.

Corolario 2. *Sea* V *un espacio con producto interior dimensionalmente finito y sea* W *un subespacio de* V. *Entonces* $\dim(W) + \dim(W^\perp) = \dim(V)$.

Ejemplo 17. Sea $V = F^3$ y $W = L(\{e_1, e_2\})$. Entonces $x = (a, b, c) \in W^\perp$ si y sólo si $0 = (x, e_1) = a$ y $0 = (x, e_2) = b$. Así, $x = (0, 0, c)$ y por lo tanto $W^\perp = L(\{e_3\})$. Se puede deducir el mismo resultado simplemente notando del Corolario 2 que $\dim(W^\perp) = 1$ y que el vector e_3 es ortogonal tanto a e_1 como a e_2.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.
   - (a) El proceso de ortogonalización de Gram-Schmidt nos permite construir un conjunto ortonormal a partir de un conjunto arbitrario de vectores.
   - (b) Todo espacio dimensionalmente finito con producto interior posee una base ortonormal.
   - (c) El complemento ortogonal de cualquier conjunto es un subespacio.

(d) Si $\beta = \{x_1, \ldots, x_n\}$ es una base para un espacio con producto interior V, entonces, para cualquier $x \in$ V, los escalares (x, x_i) $(i = 1, \ldots, n)$ son los coeficientes de Fourier de x.

(e) Para cualquier subespacio W de un espacio con producto interior dimensionalmente finito V, tenemos que $\mathsf{V} = \mathsf{W} \oplus \mathsf{W}^{\perp}$.

(f) Una base ortonormal debe ser una base ordenada.

2. En cada uno de los incisos siguientes aplicar el proceso de Gram-Schmidt al subconjunto dado S del espacio con producto interior V. Entonces encontrar una base ortonormal β para V y calcular los coeficientes de Fourier para el vector dado relativos a β. Utilizar, finalmente, el Teorema 7.5 para verificar el resultado.

(a) $\mathsf{V} = \mathsf{R}^3$, $S = \{(1, 0, 1,), (0, 1, 1), (1, 3, 3)\}$ y $x = (1, 1, 2)$

(b) $\mathsf{V} = \mathsf{R}^3$, $S = \{(1, 1, 1), (0, 1, 1), (0, 0, 1)\}$ y $x = (1, 0, 1)$

(c) $\mathsf{V} = \mathsf{P}_2(R)$ con el producto interior $(f, g) = \displaystyle\int_0^1 f(t)g(t)dt$,
$S = \{1, x, x^2\}$ y $f(x) = 1 + x$

(d) $\mathsf{V} = \mathsf{C}^3$, $S = \{(1, i, 0), (1 - i, 2, 4i)\}$ y $x = \{(i, 2 + 3i, 1)\}$

3. Sean $\mathsf{V} = \mathsf{R}^2$ y

$$\beta = \left\{\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right)\right\}.$$

Encontrar los coeficientes de Fourier para (3, 4) relativos a β.

4. Sean $\mathsf{V} = \mathsf{C}^3$ y $S = \{(1, 0, i), (1, 2, 1)\}$. Calcular $S^{\perp}$.

5. Sean $\mathsf{V} = \mathsf{R}^3$ y $S = \{x_0\}$ donde $x_0 \neq 0$. Describir geométricamente a $S^{\perp}$. Si $\{x_1, x_2\} = S_0$ es linealmente independiente, describir geométricamente a $S_0^{\perp}$.

6. Sea V un espacio con producto interior, y sea W un subespacio dimensionalmente finito de V. Si $x \notin$ W, demostrar que existe $y \in$ V tal que $y \in \mathsf{W}^{\perp}$ pero con $(x, y) \neq 0$. *Sugerencia:* Utilizar el Corolario 1 del Teorema 7.6.

7. Demostrar que si $\{y_1, \ldots, y_n\}$ es un conjunto ortogonal de vectores no nulos, entonces los vectores $\{x_1, \ldots, x_n\}$ derivados del proceso de Gram-Schmidt satisfacen a $x_i = y_i$ para $i = 1, \ldots, n$. *Sugerencia:* Utilizar inducción.

8. Sea $\mathsf{V} = \mathsf{C}^3$ con el producto interior ordinario, y sea $\mathsf{W} = L(\{(i, 0, 1)\})$. Encontrar bases ortonormales para W y $\mathsf{W}^{\perp}$.

9. Sea W un subespacio dimensionalmente finito de un espacio con producto interior V. Demostrar que existe una proyección T en W tal que $\mathsf{N}(\mathsf{T}) = \mathsf{W}^{\perp}$.

Además, demostrar que $||T(x)|| \leq ||x||$ para toda $x \in V$. *Sugerencia:* Emplear el Ejercicio 10 de la Sección 7.1.

10. Sea A una matriz de $n \times n$ con elementos complejos tales que los renglones de A forman un conjunto ortonormal. Demostrar que $AA^* = I$.

11. Sean W_1 y W_2 subespacios de un espacio con producto interior dimensionalmente finito. Demostrar que $(W_1 + W_2)^\perp = W_1^\perp \cap W_2^\perp$ y $(W_1 \cap W_2)^\perp = W_1^\perp + W_2^\perp$.

12.* Sea V un espacio con producto interior y sean S y S_0 subconjuntos de V. Demostrar los siguientes incisos:

(a) $S_0 \subseteq S$ implica que $S^\perp \subseteq S_0^\perp$.

(b) $S \subseteq (S^\perp)^\perp$, y entonces $L(S) \subseteq (S^\perp)^\perp$.

(c) Si W es un subespacio dimensionalmente finito de V, entonces $W = (W^\perp)^\perp$. *Sugerencia:* Utilizar el Ejercicio 6.

13. *Identidad de Parseval.* Sea $\{x_1, \ldots, x_n\}$ una base ortonormal para V. Demostrar, para cualesquiera $x, y \in V$, que

$$(x, y) = \sum_{i=1}^{n} (x, x_i)\overline{(y, x_i)}.$$

14. Sea V un espacio con producto interior y sea $S = \{x_1, \ldots, x_n\}$ cualquier subconjunto ortonormal de V. Demostrar que para cualquier x en V tenemos que

$$||x||^2 \geq \sum_{i=1}^{n} |(x, x_i)|^2.$$

Esta desigualdad se llama *desigualdad de Bessel. Sugerencia:* Aplicar el Corolario 1 del Teorema 7.6 a $x \in V$ y $W = L(S)$. Luego emplear el Ejercicio 10 de la Sección 7.1.

15. Sea T un operador lineal en un espacio con producto interior dimensionalmente finito V. Si $(T(x), y) = 0$ para toda $x, y \in V$, demostrar que $T = T_0$. De hecho, demostrar este resultado si la igualdad se cumple para toda x e y en alguna base para V.

16. Sea $V = C([-1, 1])$. Supóngase que W_e y W_o son los subespacios de V formados por las funciones pares e impares, respectivamente. Demostrar que $W_e^\perp = W_o$ si el producto interior en V es

$$(f, g) = \int_{-1}^{1} f(t)g(t)dt.$$

7.3 EL ADJUNTO DE UN OPERADOR LINEAL

En la Sección 7.1 definimos a la transpuesta conjugada A^* de una matriz A. Para un operador lineal T en un espacio con producto interior V, definiremos ahora un operador lineal relacionado en V llamado el "adjunto" de T, cuya matriz es $[T]^*_\beta$, donde β es cualquier base ortonormal de V. La analogía entre la conjugación compleja de números complejos y los adjuntos de los operadores lineales pronto se hará aparente. Primero, sin embargo, necesitamos un resultado preliminar.

Sea V un espacio con producto interior y sea $y \in V$. La función $g\colon V \to F$ definida mediante $g(x) = (x, y)$ para toda $x \in V$ es claramente lineal. Más importante es el hecho de que si V es dimensionalmente finito, toda transformación lineal de V en F es de esta forma.

Teorema 7.7. *Sea* V *un espacio con producto interior dimensionalmente finito sobre* F, *y sea* $g\colon V \to F$ *una transformación lineal. Entonces existe un vector único* $y \in V$ *tal que* $g(x) = (x, y)$ *para toda* $x \in V$.

DEMOSTRACIÓN. Sea β una base ortonormal para V, digamos $\beta = \{x_1, \ldots, x_n\}$, y sea

$$y = \sum_{i=1}^{n} \overline{g(x_i)} x_i.$$

Si definimos a $h\colon V \to F$ mediante $h(x) = (x, y)$, entonces h es claramente lineal. Ahora bien, para $1 \le j \le n$ tenemos

$$h(x_j) = (x_j, y) = \left(x_j, \sum_{i=1}^{n} \overline{g(x_i)} x_i\right) = \sum_{i=1}^{n} g(x_i)(x_j, x_i)$$

$$= \sum_{i=1}^{n} g(x_i)\delta_{ji} = g(x_j).$$

Como g y h coinciden en β, tenemos, de acuerdo con el corolario del Teorema 2.7, que $g = h$.

Para demostrar que y es única, supóngase que $g(x) = (x, y')$ para toda x. Entonces $(x, y) = (x, y')$ para toda x y entonces por el Teorema 7.1 tenemos que $y = y'$. ■

Ejemplo 18. Defínase a $g\colon R^2 \to R$ mediante $g(a_1, a_2) = 2a_1 + a_2$; evidentemente g es una transformación lineal. Sea $\beta = \{e_1, e_2\}$ y, como en la demostración del Teorema 7.7, sea $y = g(e_1)e_1 + g(e_2)e_2 = 2e_1 + e_2 = (2, 1)$. Entonces $g(a_1, a_2) = ((a_1, a_2), (2, 1)) = 2a_1 + a_2$.

Teorema 7.8. *Sea* V *un espacio con producto interior dimensionalmente finito y sea* T *un operador lineal en* V. *Entonces existe un operador lineal único* T^* *en* V *tal que* $(T(x), y) = (x, T^*(y))$ *para toda* $x, y \in V$.

DEMOSTRACIÓN. Sea $y \in V$. Defínase a $g: V \to F$ mediante $g(x) = (T(x), y)$ para toda $x \in V$. Primero demostraremos que g es lineal. Sean $x_1, x_2 \in V$ y $c \in F$. Entonces $g(cx_1 + x_2) = (T(cx_1 + x_2), y) = (cT(x_1) + T(x_2), y) = c(T(x_1), y) + (T(x_2), y) = cg(x_1) + g(x_2)$. Por lo tanto, g es lineal.

Ahora podemos emplear el Teorema 7.7 para obtener un vector único $y' \in V$ tal que $g(x) = (x, y')$; esto es $(T(x), y) = (x, y')$, para toda $x \in V$. Definiendo a $T^*: V \to V$ mediante $T^*(y) = y'$, tenemos que $(T(x), y) = (x, T^*(y))$.

Para demostrar que T^* es lineal, sean $y_1, y_2 \in V$ y $c \in F$. Entonces para cualquier $x \in V$ tenemos

$$\begin{aligned}(x, T^*(cy_1 + y_2)) &= (T(x), cy_1 + y_2)\\ &= \bar{c}(T(x), y_1) + (T(x), y_2)\\ &= \bar{c}(x, T^*(y_1)) + (x, T^*(y_2))\\ &= (x, cT^*(y_1) + T^*(y_2)).\end{aligned}$$

Como x es arbitraria, tenemos que $T^*(cy_1 + y_2) = cT^*(y_1) + T^*(y_2)$ de acuerdo con el Teorema 7.1(d).

Finalmente, sólo nos queda demostrar que T^* es única. Supóngase que $U: V \to V$ es lineal y satisface a $(T(x), y) = (x, U(y))$ para toda $x, y \in V$. Entonces $(x, T^*(y)) = (x, U(y))$ para toda $x, y \in V$ y finalmente $T^* = U$. ■

El operador lineal T^* descrito en el Teorema 7.8 se llama *adjunto* del operador T. El símbolo T^* se lee "T asterisco".

Luego, T^* es el único operador en V que satisface a $(T(x), y) = (x, T^*(y))$ para toda $x, y \in V$. Nótese que también

$$(x, T(y)) = \overline{(T(y), x)} = \overline{(y, T^*(x))} = (T^*(x), y)$$

y así $(x, T(y)) = (T^*(x), y)$ para toda $x, y \in V$. Podemos ver estas ecuaciones simbólicamente como que añadimos un * a T cuando cambiamos su posición dentro del símbolo de producto interior.

En el caso dimensionalmente infinito, el adjunto de un operador lineal T puede definirse como el operador lineal T^* que satisface a $(T(x), y) = (x, T^*(y))$ para toda $x, y \in V$. La unicidad de T^* se deducirá como anteriormente. Sin embargo, no se garantiza la existencia de un adjunto. El lector deberá observar la necesidad de la hipótesis de dimensionalidad finita en la demostración del Teorema 7.7. Muchos de los teoremas que demostraremos sobre adjuntas son, sin embargo, independientes de la dimensión de V. *Entonces, para el resto de este capítulo, adoptaremos para los ejercicios la convención de que, una referencia al adjunto de un operador lineal en un espacio con producto interior dimensionalmente infinito, presupone que tal adjunto existe.*

Un resultado útil para obtener adjuntos es el Teorema 7.9 que sigue.

Teorema 7.9. *Sea V un espacio con producto interior dimensionalmente finito, y sea β una base ortonormal para V. Si T es un operador lineal en V, entonces*

$$[T^*]_\beta = [T]_\beta^*.$$

DEMOSTRACIÓN. Sean $A = [T]_\beta$, $B = [T^*]_\beta$ y $\beta = \{x_1, \ldots, x_n\}$. Entonces, del corolario del Teorema 7.5, tenemos que

$$B_{ij} = (T^*(x_j), x_i) = \overline{(x_i, T^*(x_j))}$$
$$= \overline{(T(x_i), x_j)} = \overline{A_{ji}} = (A^*)_{ij}.$$

Por lo tanto $B = A^*$. ■

Corolario. *Sea A una matriz de $n \times n$. Entonces* $L_{A^*} = (L_A)^*$.

DEMOSTRACIÓN. Si β es la base ordenada estándar para F^n entonces, en virtud del Teorema 2.17, tenemos que $[L_A]_\beta = A$. Por lo tanto, $[(L_A)^*]_\beta = [L_A]_\beta^* = A^* = [L_{A^*}]_\beta$, y así $(L_A)^* = L_{A^*}$. ■

Como aplicación de lo anterior, calcularemos el adjunto de un operador lineal específico.

Ejemplo 19. Sea $T: C^2 \to C^2$ definido mediante $T(a_1, a_2) = (2ia_1 + 3a_2, a_1 - a_2)$. Si β es la base ordenada estándar para C^2, entonces

$$[T]_\beta = \begin{pmatrix} 2i & 3 \\ 1 & -1 \end{pmatrix}.$$

Luego

$$[T^*]_\beta = [T]_\beta^* = \begin{pmatrix} -2i & 1 \\ 3 & -1 \end{pmatrix}.$$

Por lo tanto

$$T^*(a_1, a_2) = (-2ia_1 + a_2, 3a_1 - a_2).$$

El siguiente teorema demuestra la analogía entre los complejos conjugados de los números complejos y los adjuntos de los operadores lineales.

Teorema 7.10. *Sea V un espacio con producto interior dimensionalmente finito y sean T y U operadores lineales en V. Entonces*

(a) $(T + U)^* = T^* + U^*$.
(b) $(cT)^* = \bar{c}T^*$ *para cualquier* $c \in F$.
(c) $(TU)^* = U^*T^*$.
(d) $T^{**} = T$.
(e) $I^* = I$.

DEMOSTRACIÓN. Demostraremos (a) y (d); el resto se demuestra de la misma manera. Sea $x, y \in V$.

Como

$$\begin{aligned}(x, (T+U)^*(y)) &= ((T+U)(x), y) = (T(x)+U(x), y)\\ &= (T(x), y) + (U(x), y) = (x, T^*(y)) + (x, U^*(y))\\ &= (x, T^*(y) + U^*(y)) = (x, (T^*+U^*)(y)),\end{aligned}$$

se sigue el inciso (a).

De la misma manera, como

$$\begin{aligned}(x, T(y)) &= (T^*(x), y)\\ &= (x, T^{**}(y)),\end{aligned}$$

se obtiene (d). ■

La misma demostración opera en el caso dimensionalmente infinito, siempre y cuando se suponga la existencia de T^* y U^*.

Corolario. *Sean* A *y* B *matrices de* $n \times n$. *Entonces*

(a) $(A+B)^* = A^* + B^*$.
(b) $(\bar{c}A)^* = \bar{c}A^*$ *para toda* $c \in F$.
(c) $(AB)^* = B^*A^*$.
(d) $A^{**} = A$.
(e) $I^* = I$.

DEMOSTRACIÓN. Demostraremos únicamente el inciso (c), los incisos restantes se pueden demostrar de manera semejante.

Como $L_{(AB)^*} = (L_{AB})^* = (L_A L_B)^* = (L_B)^*(L_A)^* = L_{B^*}L_{A^*} = L_{B^*A^*}$, tenemos que $(AB)^* = B^*A^*$. ■

En la demostración anterior nos apoyamos en el corolario del Teorema 7.9. Una demostración alternativa se daría acudiendo directamente a la definición de transpuesta conjugada. (Ver Ejercicio 5.)

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas. Considérese que los espacios con producto interior subyacentes son dimensionalmente finitos.

   (a) Todo operador lineal tiene un adjunto.
   (b) Todo operador lineal en V tiene la forma $x \to (x, y)$ para alguna $y \in V$.
   (c) Para todo operador lineal T en V y toda base β de V, $[T^*]_\beta = ([T]_\beta)^*$.

(d) El adjunto de un operador lineal es siempre único.
(e) Para operadores cualesquiera T y U, y escalares a y b,

$$(a\mathsf{T} + b\mathsf{U})^* = a\mathsf{T}^* + b\mathsf{U}^*.$$

(f) Para cualquier matriz A de $n \times n$, $(\mathsf{L}_A)^* = \mathsf{L}_{A^*}$.
(g) Para cualquier operador T, $(\mathsf{T}^*)^* = \mathsf{T}$.

2. Para cada uno de los siguientes espacios con producto interior V (sobre F) y transformaciones lineales g: $\mathsf{V} \to F$, encontrar un vector y tal que $\mathsf{g}(x) = (x, y)$ para toda $x \in \mathsf{V}$.

(a) $\mathsf{V} = \mathsf{R}^3$, $\mathsf{g}(a_1, a_2, a_3) = a_1 - 2a_2 + 4a_3$
(b) $\mathsf{V} = \mathsf{C}^2$, $\mathsf{g}(z_1, z_2) = z_1 - 2z_2$

(c) $\mathsf{V} = \mathsf{P}_2(R)$ con $(f, h) = \int_0^1 f(t)h(t)dt$, $\quad \mathsf{g}(f) = f(0) + f'(1)$

3. Para cada uno de los siguientes espacios con producto interior V y operadores lineales T en V, evaluar T* en el elemento dado de V.

(a) $\mathsf{V} = \mathsf{R}^2$, $\quad \mathsf{T}(a, b) = (2a + b, a - 3b)$, $\quad x = (3, 5)$
(b) $\mathsf{V} = \mathsf{C}^2$, $\quad \mathsf{T}(z_1, z_2) = (2z_1 + iz_2, (1 - i)z_1)$,

$$x = (3 - i, 1 + 2i)$$

(c) $\mathsf{V} = \mathsf{P}_2(R)$ con

$$(f, g) = \int_0^1 f(t)g(t)dt, \qquad \mathsf{T}(f) = f' + 3f,$$

$$f(x) = 4 - x + 3x^2$$

4. Completar la demostración del Teorema 7.10.

5. Completar la demostración del corolario del Teorema 7.10 de dos maneras. Primero emplear el Teorema 7.10 tal como en la demostración de (c). Luego emplear la definición matricial de A^*.

6. Sea T un operador lineal en un espacio con producto interior V. Sean $\mathsf{U}_1 = \mathsf{T} + \mathsf{T}^*$ y $\mathsf{U}_2 = \mathsf{T}\mathsf{T}^*$. Demostrar que $\mathsf{U}_1 = \mathsf{U}_1^*$ y $\mathsf{U}_2 = \mathsf{U}_2^*$.

7. Dar un ejemplo de un operador lineal T en un espacio con producto interior V tal que $\mathsf{N}(\mathsf{T}) \neq \mathsf{N}(\mathsf{T}^*)$.

8. Sea V un espacio con producto interior dimensionalmente finito, y sea T un operador lineal en V. Demostrar que si T es invertible, entonces T* es invertible y $(\mathsf{T}^*)^{-1} = (\mathsf{T}^{-1})^*$.

9. Demostrar que si $\mathsf{V} = \mathsf{W} \oplus \mathsf{W}^\perp$ y T es la proyección en W con $\mathsf{N}(\mathsf{T}) = \mathsf{W}^\perp$, entonces $\mathsf{T} = \mathsf{T}^*$.

10. Sea T un operador lineal en un espacio con producto interior V. Demostrar que $||T(x)|| = ||x||$ para toda $x \in V$ si y sólo si $(T(x), T(y)) = (x, y)$ para toda $x, y \in V$. *Sugerencia:* Emplear el Ejercicio 20 de la Sección 7.1.

11. Para un operador lineal T en un espacio con producto interior V, demostrar que $T^*T = T_0$ implica que $T = T_0$. ¿Es cierto el mismo resultado si suponemos que $TT^* = T_0$?

12.* Sea V un espacio con producto interior dimensionalmente finito y, sea T un operador lineal en V. Demostrar que $R(T^*) = N(T)^{\perp}$. *Sugerencia:* Demostrar que $R(T^*)^{\perp} = N(T)$, y luego utilizar el Ejercicio 12(c) de la Sección 7.2.

13. Sea T un operador lineal en un espacio con producto interior dimensionalmente finito V. Demostrar los siguientes incisos.

(a) $N(T^*T) = N(T)$. Deducir que $\text{rango}(T^*T) = \text{rango}(T)$.
(b) $\text{rango}(T) = \text{rango}(T^*)$. Deducir de (a) que $\text{rango}(TT^*) = \text{rango}(T)$.
(c) Para cualquier matriz A de $n \times n$, $\text{rango}(A^*A) = \text{rango}(AA^*) = \text{rango}(A)$.

14. Sea V un espacio con producto interior y sea $y, z \in V$. Defínase a $T: V \to V$ mediante $T(x) = (x, y)z$ para toda $x \in V$. Primero demostrar que T es lineal. Luego demostrar que T^* existe y definirla explícitamente.

15. Sea $T: V \to W$ una transformación lineal entre los espacios con producto interior dimensionalmente finitos V y W.

(a) Demostrar que existe una transformación lineal única $T^*: W \to V$ tal que $(T(x), y) = (x, T^*(y))$ para toda $x \in V$, $y \in W$.
(b) Sean β y γ las bases ortonormales respectivas para V y W. Demostrar que $[T^*]_{\gamma}^{\beta} = ([T]_{\beta}^{\gamma})^*$.

16. Sea A una matriz de $n \times n$. Demostrar que $\det(A^*) = \overline{\det(A)}$.

7.4* LA TEORIA ESPECIAL DE LA RELATIVIDAD DE EINSTEIN

Como resultado de los experimentos físicos realizados durante la segunda mitad del siglo diecinueve (de una manera más sobresaliente, el experimento de Michelson-Morley de 1887), los físicos concluyeron que los *resultados obtenidos en la medición de la velocidad de la luz son independientes de la velocidad del instrumento utilizado para medirla.* Por ejemplo, supóngase que estando en la Tierra un experimentador mide la velocidad de la luz emitida por el Sol y encuentra que es de 300 000 kilómetros por segundo. Ahora supóngase que el experimentador coloca

el equipo de medición en una nave espacial y abandona la Tierra viajando a 160 000 kilómetros por segundo alejándose del Sol. Una repetición del mismo experimento desde la nave espacial arrojaría el mismo resultado: ¡La luz está viajando a 300 000 kilómetros por segundo relativa a la nave espacial y no a 140 000 kilómetros por segundo como era de esperarse.

Esta revelación condujo a una nueva manera de relacionar los sistemas coordenados empleados para ubicar a los eventos en el espacio-tiempo. El resultado fue la teoría especial de la relatividad de Albert Einstein. Desarrollaremos la esencia de la teoría de Einstein desde el punto de vista del álgebra lineal.

El problema básico es comparar dos distintos sistemas de coordenadas sin aceleración, que están en movimiento relativo uno con respecto al otro bajo la suposición de que la velocidad de la luz es la misma, medida en ambos sistemas. Supóngase que nos dan dos sistemas coordenados inerciales (sin aceleración) S y S' en un espacio de tres dimensiones (R^3) y tales que S' se desplaza a una velocidad constante en relación con S, medida a partir de S (ver la Fig. 7.2). Para simplificar las cosas, supongamos que:

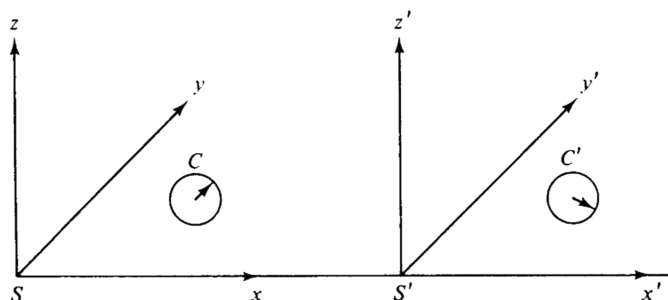


figura 7.2

1. Los ejes correspondientes de S y S' (x y x', y e y', z y z') son paralelos y el origen de S' se desplaza en la dirección positiva del eje x de S a una velocidad constante $v > 0$ relativa a S.
2. Se colocan dos relojes C y C' en el espacio —el primero estacionario relativo al sistema de coordenadas S y el segundo estacionario relativo al sistema de coordenadas S'. Estos relojes están diseñados para dar como lecturas números reales en unidades de tiempo (segundos). Se calibran los relojes de manera que en el instante en que los orígenes de S y S' coincidan, ambos relojes den la lectura cero.
3. Nuestra unidad de longitud será el *segundo luz* (la distancia que recorre la luz en un segundo) y nuestra unidad de tiempo será el segundo. Nótese que con respecto a estas unidades la velocidad de la luz es de un segundo luz por segundo.

Dado un evento cualquiera (cualquier cosa cuya posición y tiempo de ocurrencia pueda ser descrito) le podemos asignar un conjunto de coordenadas de "espacio-tiempo". Por ejemplo, si p es un evento que ocurre en una posición

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

relativa a S en un tiempo t leído en el reloj C, podemos asignar a p el conjunto de coordenadas

$$\begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix}$$

Esta cuarteta ordenada se denomina *las coordenadas espacio-tiempo* de p relativas a S y a C. De la misma manera p tiene un conjunto de coordenadas espacio-tiempo

$$\begin{pmatrix} x' \\ y' \\ z' \\ t' \end{pmatrix}$$

relativas a S' y a C'.

Podemos definir un mapeo T_v: $\mathsf{R}^4 \to \mathsf{R}^4$ (que depende de la velocidad v) como consecuencia de lo anterior tal que, para cualquier conjunto de coordenadas de espacio-tiempo

$$\begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix}$$

que miden un evento con respecto a S y a C,

$$\mathsf{T}_v \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{pmatrix} x' \\ y' \\ z' \\ t' \end{pmatrix}$$

es el conjunto de coordenadas espacio-tiempo de este evento con respecto a S' y a C'. Es evidente que T_v es uno-a-uno y sobreyectivo.

Einstein hizo ciertas suposiciones sobre T_v que condujeron a su teoría especial de la relatividad. Formularemos un conjunto equivalente de suposiciones.

Axiomas de la teoría especial de la relatividad

R_1: La velocidad de cualquier haz de luz, al ser medida en cualquiera de los sistemas coordenados utilizando un reloj estacionario relativo al mismo sistema, es 1.

R_2: El mapeo T_v: $R^4 \to R^4$ es lineal.

R_3: Para cualquier

$$\begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} \in R^4,$$

si

$$T_v \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{pmatrix} x' \\ y' \\ z' \\ t' \end{pmatrix},$$

entonces $y' = y$ y $z' = z$.

R_4: Para

$$T_v \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{pmatrix} x' \\ y' \\ z' \\ t' \end{pmatrix},$$

x' y t' son independientes de y y z; esto es, si

$$T_v \begin{pmatrix} x \\ y_1 \\ z_1 \\ t \end{pmatrix} = \begin{pmatrix} x' \\ y' \\ z' \\ t' \end{pmatrix} \quad \text{y} \quad T_v \begin{pmatrix} x \\ y_2 \\ z_2 \\ t \end{pmatrix} = \begin{pmatrix} x'' \\ y'' \\ z'' \\ t'' \end{pmatrix},$$

entonces $x'' = x'$ y $t'' = t'$.

R_5: El origen de S se desplaza en la dirección negativa del eje X' de S' a una velocidad constante $-v < 0$ medida desde S'.

Como veremos, estos 5 axiomas (R_1, R_2, R_3, R_4 y R_5) caracterizan completamente a T_v. El operador T_v se llama *transformación de Lorentz en la dirección* x. Pretendemos calcular T_v y utilizarla para estudiar los curiosos fenómenos de la contracción del tiempo.

Teorema 7.11. *En* R^4

(a) $T_v(e_i) = e_i$ *para* $i = 2, 3$.

(b) $L(\{e_2, e_3\})$ *es* T_v*-invariante.*

(c) $L(\{e_1, e_4\})$ *es* T_v*-invariante.*

(d) $L(\{e_2, e_3\})$ *y* $L(\{e_1, e_4\})$ *son* T_v^**-invariantes.*

(e) $T_v^*(e_i) = e_i$ *para* $i = 2, 3$.

DEMOSTRACIÓN.

(a) Por el axioma R_2

$$T_v\begin{pmatrix}0\\0\\0\\0\end{pmatrix} = \begin{pmatrix}0\\0\\0\\0\end{pmatrix},$$

y por lo tanto, por el axioma R_4, las coordenadas primera y cuarta de

$$T_v\begin{pmatrix}0\\a\\b\\0\end{pmatrix}$$

son iguales a cero para cualquier $a, b \in R$. Luego, por el axioma R_3

$$T_v\begin{pmatrix}0\\1\\0\\0\end{pmatrix} = \begin{pmatrix}0\\1\\0\\0\end{pmatrix} \quad \text{y} \quad T_v\begin{pmatrix}0\\0\\1\\0\end{pmatrix} = \begin{pmatrix}0\\0\\1\\0\end{pmatrix}.$$

Las demostraciones de los incisos (b), (c) y (d) se dejan como ejercicios.

(e) Para cualquier $j \neq 2$, en virtud de (a) y (c), $(T_v^*(e_2), e_j) = (e_2, T_v(e_j)) = 0$; para $j = 2$, por (a), $(T_v^*(e_2), e_j) = (e_2, T_v(e_2)) = (e_2, e_2) = 1$. Concluimos que $T_v^*(e_2)$ es un múltiplo de e_2, o sea que $T_v^*(e_2) = \lambda e_2$ para alguna $\lambda \in R$. Entonces $1 = (e_2, e_2) = (e_2, T_v(e_2)) = (T_v^*(e_2), e_2) = (\lambda e_2, e_2) = \lambda$ y, por lo tanto, $T_v^*(e_2) = e_2$. De la misma manera $T_v^*(e_3) = e_3$. ■

Supóngase que en el instante en el que los orígenes de S y S' coinciden se emite un destello luminoso desde su origen común. Cuando este evento se mide relativo a S y C o relativo a S' y C' tiene coordenadas espacio-tiempo

$$\begin{pmatrix}0\\0\\0\\0\end{pmatrix}.$$

Sea P el conjunto de todos los eventos cuyas coordenadas espacio-tiempo

$$\begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix}$$

relativas a S y C son tales que el destello se observa en el punto de coordenadas

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

(medidas con respecto a S) en el instante t (medido en C). Permítasenos caracterizar a P en términos de x, y, z y t. Como la velocidad de la luz es 1, en cualquier instante $t \geq 0$ el destello se observa desde cualquier punto cuya distancia al origen de S (medida a partir de S) sea $t \cdot 1 = t$. Estos son justamente los puntos que se localizan sobre la superficie de la esfera de radio t con centro en el origen Las coordenadas (relativas a S) de tales puntos satisfacen la ecuación $x^2 + y^2 + z^2 = t^2$. Por lo tanto, un evento está en P si y sólo si sus coordenadas espacio-tiempo relativas a S y a C

$$\begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} \qquad (t \geq 0)$$

satisfacen la ecuación $x^2 + y^2 + z^2 - t^2 = 0$. En virtud del axioma R_1 podemos caracterizar a P en términos de las coordenadas espacio-tiempo relativas a S' y a C' de la misma manera: un evento está en P si y sólo si sus coordenadas espacio-tiempo relativas a S' y a C'

$$\begin{pmatrix} x' \\ y' \\ z' \\ t' \end{pmatrix} \qquad (t' \geq 0)$$

satisface la ecuación $(x')^2 + (y')^2 + (z')^2 - (t')^2 = 0$.

Sea

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Teorema 7.12. *Para cualquier* $w \in R^4$, *si* $(L_A(w), w) = 0$, *entonces* $(T_v^* L_A T_v(w), w) = 0$.

DEMOSTRACIÓN. Sea

$$w = \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} \in R^4,$$

y supóngase que $(L_A(w), w) = 0$.

CASO 1. $t \geq 0$. Como $(L_A(w), w) = x^2 + y^2 + z^2 - t^2$, w es el conjunto de coordenadas de un evento de P relativos a S y a C. Como

$$\begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} x' \\ y' \\ z' \\ t' \end{pmatrix}$$

son las coordenadas espacio-tiempo del mismo evento relativas a S' y a C', la discusión que precede al Teorema 7.12 da

$$(x')^2 + (y')^2 + (z')^2 - (t')^2 = 0.$$

Luego entonces, $(T_v^* L_A T_v(w), w) = (L_A T_v(w), T_v(w)) = (x')^2 + (y')^2 + (z')^2 - (t')^2 = 0$, y se obtiene la conclusión.

CASO 2. $t < 0$. La demostración se obtiene al aplicar el Caso 1 a $-w$. ∎

Procedamos ahora a deducir información acerca de T_v. Sean

$$w_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad \text{y} \quad w_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}.$$

Por el Ejercicio 3, $\{w_1, w_2\}$ es una base ortogonal para $L(\{e_1, e_4\})$, y $L(\{e_1, e_4\})$ es $T_v^* L_A T_v$-invariante. El siguiente resultado nos dice aún más.

Teorema 7.13. *Existen escalares* a *y* b *no nulos tales que*

(a) $T_v^* L_A T_v(w_1) = aw_2$,
(b) $T_v^* L_A T_v(w_2) = bw_1$.

DEMOSTRACIÓN.

(a) De acuerdo con el Teorema 7.12, $(L_A(w_1), w_1) = 0$, $(T_v^* L_A T_v(w_1), w_1) = 0$. Entonces, $T_v^* L_A T_v(w_1)$ es ortogonal a w_1. Como $L(\{e_1,$

$e_4\}) = L(\{w_1, w_2\})$ es $\mathsf{T}_v^*\mathsf{L}_A\mathsf{T}_v$-invariante, $\mathsf{T}_v^*\mathsf{L}_A\mathsf{T}_v(w_1)$ debe pertenecer a este conjunto. Pero $\{w_1, w_2\}$ es una base ortogonal para este subespacio, y entonces $\mathsf{T}_v^*\mathsf{L}_A\mathsf{T}_v(w_1)$ debe ser un múltiplo de w_2. Así, $\mathsf{T}_v^*\mathsf{L}_A\mathsf{T}_v(w_1) = aw_2$ para algún escalar a. Como T_v y A son invertibles, también $\mathsf{T}_v^*\mathsf{L}_A\mathsf{T}_v$ lo es. Luego $a \neq 0$, demostrando así (a). La demostración de (b) es semejante. ■

Corolario. *Sea* $\mathrm{B}_v = [\mathsf{T}_v]_\beta$, *donde* β *es la base ordenada estándar para* R^4. *Entonces*

(a) $\mathrm{B}_v^*A\mathrm{B}_v = A$.
(b) $\mathsf{T}_v^*\mathsf{L}_A\mathsf{T}_v = \mathsf{L}_A$.

Dejaremos como ejercicio la prueba del corolario. Para algunas sugerencias, véase el Ejercicio 4.

Consideremos ahora la situación cuando ha transcurrido un segundo desde que los orígenes S y S' coincidieran medido por el reloj C. Como el origen de S' se desplaza a lo largo del eje x con una velocidad v medida en S, sus coordenadas espacio-tiempo relativas a S y C son

$$\begin{pmatrix} v \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

De la misma manera, las coordenadas espacio-tiempo para el origen de S' relativas a S' y a C' deben ser

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ t' \end{pmatrix}$$

para algún $t' > 0$. Entonces tenemos que

$$\mathsf{T}_v\begin{pmatrix} v \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ t' \end{pmatrix} \quad \text{para algún } t' > 0. \tag{3}$$

Por el corolario del Teorema 7.13

$$\left(\mathsf{T}_v^*\mathsf{L}_A\mathsf{T}_v\begin{pmatrix} v \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} v \\ 0 \\ 0 \\ 1 \end{pmatrix}\right) = \left(\mathsf{L}_A\begin{pmatrix} v \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} v \\ 0 \\ 0 \\ 1 \end{pmatrix}\right) = v^2 - 1. \tag{4}$$

Pero también

$$\left(\mathsf{T}_v^*\mathsf{L}_A\mathsf{T}_v\begin{pmatrix} v \\ 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} v \\ 0 \\ 0 \\ 1 \end{pmatrix}\right) = \left(\mathsf{L}_A\mathsf{T}_v\begin{pmatrix} v \\ 0 \\ 0 \\ 1 \end{pmatrix}, \mathsf{T}_v\begin{pmatrix} v \\ 0 \\ 0 \\ 1 \end{pmatrix}\right) = \left(\mathsf{L}_A\begin{pmatrix} 0 \\ 0 \\ 0 \\ t' \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ t' \end{pmatrix}\right) = -(t')^2. \tag{5}$$

Combinando las ecuaciones (4) y (5), concluimos que

$$v^2 - 1 = -(t')^2, \quad \text{o bien} \quad t' = \sqrt{1 - v^2}. \tag{6}$$

Luego, de las ecuaciones (3) y (6), obtenemos

$$\mathsf{T}_v\begin{pmatrix} v \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \sqrt{1 - v^2} \end{pmatrix}. \tag{7}$$

Recuérdese luego que el origen de S se desplaza en la dirección negativa del eje x de S' con la velocidad constante $-v < 0$ medida desde S'. (Este hecho es el axioma R_5.) En consecuencia, un segundo después de que los orígenes de S y S' coincidieran medido con el reloj C, existe un tiempo $t' > 0$ medido en el reloj C' tal que

$$\mathsf{T}_v\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -vt' \\ 0 \\ 0 \\ t' \end{pmatrix}. \tag{8}$$

De la Ecuación (8) se obtiene, de una manera semejante a como se obtuvo la Ecuación (7), que

$$t' = \frac{1}{\sqrt{1 - v^2}}, \tag{9}$$

y, por lo tanto, de las Ecuaciones (8) y (9)

$$\mathsf{T}_v\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \dfrac{-v}{\sqrt{1 - v^2}} \\ 0 \\ 0 \\ \dfrac{1}{\sqrt{1 - v^2}} \end{pmatrix}. \tag{10}$$

El siguiente resultado se puede demostrar fácilmente utilizando las Ecuaciones (7) y (10), y el Teorema 7.11.

Teorema 7.14. *Sea β la base ordenada estándar para* R^4. *Entonces*

$$[T_v]_\beta = B_v = \begin{pmatrix} \frac{1}{\sqrt{1-v^2}} & 0 & 0 & \frac{-v}{\sqrt{1-v^2}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{-v}{\sqrt{1-v^2}} & 0 & 0 & \frac{1}{\sqrt{1-v^2}} \end{pmatrix}.$$

Contracción del tiempo

Una conclusión por demás curiosa y paradójica se deriva si aceptamos la teoría de Einstein, la de la contracción del tiempo. Supóngase que un astronauta abandona nuestro sistema solar en una nave espacial que viaja a una velocidad fija v medida con respecto a nuestro sistema solar. Se tiene de la teoría de Einstein que al final del tiempo t medido desde la Tierra, el tiempo que habrá transcurrido en la nave espacial es únicamente $t\sqrt{1-v^2}$. Para establecer este resultado, considérense los mismos sistemas de coordenadas S y S' y los relojes C y C' que estudiamos antes. Supóngase que el origen de S' coincide con la nave espacial y que el origen de S coincide con un punto en el sistema solar (estacionario con relación al Sol), de manera que los orígenes de S y S' coincidan y los relojes C y C' den una lectura de cero en el momento en el que el astronauta inicia su viaje.

Visto desde S, las coordenadas espacio-tiempo del vehículo en cualquier instante $t > 0$ medidas por C son

$$\begin{pmatrix} vt \\ 0 \\ 0 \\ t \end{pmatrix},$$

mientras que vistas desde S' las coordenadas espacio-tiempo del vehículo en cualquier instante $t' > 0$ medidas por C' son

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ t' \end{pmatrix}.$$

Pero si dos conjuntos de coordenadas de espacio-tiempo

$$\begin{pmatrix} vt \\ 0 \\ 0 \\ t \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ t' \end{pmatrix}$$

describen el mismo evento, debe tenerse que

$$T_v\begin{pmatrix} vt \\ 0 \\ 0 \\ t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ t' \end{pmatrix}.$$

Luego entonces

$$\begin{pmatrix} \frac{1}{\sqrt{1-v^2}} & 0 & 0 & \frac{-v}{\sqrt{1-v^2}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{-v}{\sqrt{1-v^2}} & 0 & 0 & \frac{1}{\sqrt{1-v^2}} \end{pmatrix}\begin{pmatrix} vt \\ 0 \\ 0 \\ t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ t' \end{pmatrix}.$$

De la ecuación anterior se tiene que

$$\frac{-v^2t}{\sqrt{1-v^2}} + \frac{t}{\sqrt{1-v^2}} = t', \quad \text{o bien} \quad t' = t\sqrt{1-v^2} \qquad (11)$$

Este es el resultado deseado.

Una consecuencia dramática de la contracción del tiempo la proporciona el Ejercicio 9 al final de esta sección.

Hagamos una consideración adicional. Supóngase que las unidades de distancia y tiempo que consideramos son unidades que se usan más comúnmente que el segundo-luz y el segundo, tales como la milla y la hora, o el kilómetro y el segundo. Sea c la velocidad de la luz en las unidades que hayamos seleccionado para la distancia y el tiempo. Se puede ver fácilmente que si un objeto viaja a una velocidad v relativa a un conjunto de unidades, entonces viaja a una velocidad v/c en unidades de segundos-luz por segundo. Así, para un conjunto cualquiera de unidades de distancia y tiempo, la Ecuación (11) se transforma en

$$t' = t\sqrt{1 - \frac{v^2}{c^2}}.$$

EJERCICIOS

1. Demostrar los incisos (b), (c) y (d) del Teorema 7.11.

2. Completar la demostración del Teorema 7.12 para el caso $t < 0$.

3. Para

$$w_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad \text{y} \quad w_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$$

demostrar que

(a) $\{w_1, w_2\}$ es una base ortogonal para $L(\{e_1, e_4\})$.
(b) $L(\{e_1, e_4\})$ es $\mathsf{T}_v^* \mathsf{L}_A \mathsf{T}_v$-invariante.

4. Demostrar el corolario del Teorema 7.13.

Sugerencias:

(a) Demostrar que

$$B_v^* A B_v = \begin{pmatrix} p & 0 & 0 & q \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -q & 0 & 0 & -p \end{pmatrix},$$

donde

$$p = \frac{a+b}{2} \quad \text{y} \quad q = \frac{a-b}{2}.$$

(b) Demostrar que $q = 0$ utilizando el hecho de que $B_v^* A B_v$ es autoadjunta.
(c) Aplicar el Teorema 7.12 a

$$w = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

para demostrar que $p = 1$.

5. Demostrar que

$$\mathsf{T}_v \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \dfrac{-v}{\sqrt{1-v^2}} \\ 0 \\ 0 \\ \dfrac{1}{\sqrt{1-v^2}} \end{pmatrix} \quad \text{[Ec. (10)].}$$

Sugerencia: Utilizar una técnica similar a la que se empleó en la obtención de la Ecuación (7).

6. Dados tres sistemas coordenados S, S' y S'' con ejes correspondientes (x, x', x''; y, y', y''; y z, z', z'') paralelos y tales que los ejes x, x' y x'' coinciden. Supóngase que S' se mueve delante de S con una velocidad $v_1 > 0$ (medida en S), S'' se mueve delante de S' con una velocidad $v_2 > 0$ (medida en S') y S'' se mueve delante de S con una velocidad $v_3 > 0$ (medida en S), y que existen tres relojes C, C' y C'' tales que C es estacionario con relación a S, C' es estacionario con relación a S' y C'' es estacionario con relación a S''. Supóngase que cuando en cualquiera de los tres relojes se mide el tiempo cero, los orígenes de S, S' y S'' coinciden. Suponiendo que $T_{v_3} = T_{v_2}T_{v_1}$ (esto es $B_{v_3} = B_{v_2}B_{v_1}$), demostrar que

$$v_3 = \frac{v_1 + v_2}{1 + v_1 v_2}.$$

Nótese que el sustituir $v_2 = 1$ en la ecuación anterior se obtiene $v_3 = 1$. Esto nos dice que la velocidad de la luz medida en S o S' es la misma. ¿Por qué nos sorprendería si esto no ocurriera?

7. Calcular $(B_v)^{-1}$. Demostrar que $(B_v)^{-1} = B_{(-v)}$. Conclúyase que si S' se mueve a una velocidad negativa v relativa a S, entonces $[T_v]_\beta = B_v$, donde B_v tiene la forma dada en el Teorema 7.14.

8. Supóngase que un astronauta abandonó la Tierra en el año de 1776 y viajó a una estrella situada a 99 años luz de la Tierra a una velocidad de 99% de la velocidad de la luz, y llegando a la estrella emprendió de inmediato el regreso a la Tierra a la misma velocidad. Considerando la teoría especial de la relatividad de Einstein, demostrar que si el astronauta tenía 20 años de edad en el momento de su partida, regresaría a la Tierra a la edad de 48.2 años en el año de 1976. Explicar la utilidad del Ejercicio 7 al resolver este problema.

9. Recuérdese la nave espacial en movimiento considerada en el estudio de la contracción del tiempo. Supóngase que el vehículo se desplaza hacia una estrella fija localizada en el eje x de S a una distancia de b unidades del origen de S. Si la nave espacial viaja hacia la estrella a una velocidad v, los habitantes de la Tierra (que permanecen "casi" estacionarios con respecto a S) calcularán que el tiempo que le toma al vehículo alcanzar la estrella es $t = b/v$. Debido al fenómeno de contracción del tiempo el astronauta percibirá un tiempo de $t' = t\sqrt{1 - v^2} = (b/v)\sqrt{1 - v^2}$. Aparece una paradoja en el hecho de que el astronauta percibe un tiempo inconsistente con una distancia b y una velocidad v. Pero la paradoja se resuelve observando que la distancia desde el sistema solar a la estrella medida por el astronauta es menor que b.

Suponiendo que los sistemas coordenados S y S' y los relojes C y C' son los descritos en la exposición de la contracción del tiempo,

(a) Discutir que en el tiempo t (medido en C) las coordenadas espacio-tiempo de la estrella relativas a S y C son

$$\begin{pmatrix} b \\ 0 \\ 0 \\ t \end{pmatrix}.$$

(b) Demostrar que en el tiempo t (medido en C) las coordenadas espacio-tiempo de la estrella relativas a S' y C' son

$$\begin{pmatrix} \dfrac{b - vt}{\sqrt{1 - v^2}} \\ 0 \\ 0 \\ \dfrac{t - bv}{\sqrt{1 - v^2}} \end{pmatrix}.$$

(c) Haciendo

$$x' = \frac{b - tv}{\sqrt{1 - v^2}} \quad \text{y} \quad t' = \frac{t - bv}{\sqrt{1 - v^2}},$$

demostrar que $x' = b\sqrt{1 - v^2} - t'v$.

Este resultado se puede interpretar como que en el tiempo t' medido por el astronauta, la distancia del astronauta a la estrella, medida por el astronauta (ver la Fig. 7.3), es

$$b\sqrt{1 - v^2} - t'v.$$

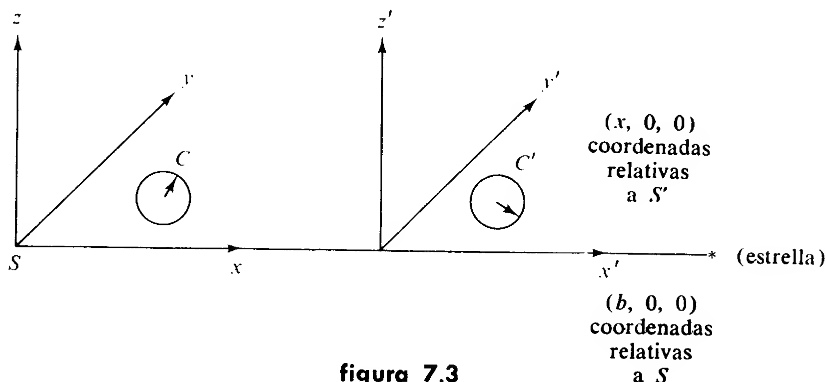


figura 7.3

(d) Conclúyase de esto que

(i) La velocidad de la nave espacial con respecto a la estrella medida por el astronauta es v.

(ii) La distancia de la Tierra a la estrella medida por el astronauta es $b\sqrt{1 - v^2}$.

Así, las distancias a lo largo de la línea de movimiento del vehículo espacial aparecen contraídas por un factor de $\sqrt{1-v^2}$.

7.5 OPERADORES NORMALES Y AUTOADJUNTOS

En esta sección demostraremos que en cualquier espacio con producto interior dimensionalmente finito existe una clase de operadores lineales totalmente determinados por sus eigenvalores. Específicamente, para un tal operador lineal T, existe una base ortonormal de eigenvectores β. Dado que las matrices que representan a T y T* en la base ordenada β son matrices diagonales, tales matrices conmutan; por lo tanto, la condición $TT^* = T^*T$ se hace necesaria para la existencia de tal base β. Demostraremos que para espacios complejos con producto interior esta condición también es suficiente.

Definiciones. *Sea* V *un espacio con producto interior, y sea* T *un operador lineal en* V. *Decimos que* T *es normal si* $TT^* = T^*T$. *Una matriz* A *de* $n \times n$ *es* normal *si* $AA^* = A^*A$.

Nótese que si β es una base ortonormal finita formada por eigenvectores de T, entonces T es normal si y sólo si $[T]_\beta$ es normal. Por supuesto, cualquier matriz diagonal es normal.

Si V no es dimensionalmente finito, entonces para que T sea normal es necesario que T* exista.

Con el objeto de construir una base ortonormal de eigenvectores para un operador normal T, debemos demostrar primero que cualquiera de estos operadores tiene al menos un eigenvector. Para este resultado requerimos del siguiente teorema.

Teorema 7.15. *Sea* V *un espacio con producto interior, y sea* T *un operador normal en* V. *Entonces*

(a) $\|T(x)\| = \|T^*(x)\|$ *para toda* $x \in V$.

(b) $T - cI$ *es normal para toda* $c \in F$.

(c) *Si* λ *es un eigenvalor de* T, *entonces* $\bar{\lambda}$ *es un eigenvalor de* T*. *De hecho,* $T(x) = \lambda x$ *implica que* $T^*(x) = \bar{\lambda}x$.

(d) *Si* λ_1 *y* λ_2 *son distintos eigenvalores de* T *con eigenvectores correspondientes* x_1 *y* x_2, *entonces* x_1 *y* x_2 *son ortogonales.*

DEMOSTRACIÓN.

(a) Para cualquier $x \in V$, tenemos que

$$\begin{aligned}\|T(x)\|^2 &= (T(x), T(x)) = (T^*T(x), x)\\ &= (TT^*(x), x) = (T^*(x), T^*(x)) = \|T^*(x)\|^2.\end{aligned}$$

La demostración de (b) se deja como ejercicio.

(c) Sea $U = T - \lambda I$ y supóngase que $T(x) = \lambda x$ para alguna $x \in V$. Entonces $U(x) = 0$, y en virtud de (a) y (b) tenemos que

$$0 = \|U(x)\| = \|U^*(x)\| = \|(T^* - \bar{\lambda}I)(x)\| = \|T^*(x) - \bar{\lambda}x\|.$$

Por lo tanto $T^*(x) = \bar{\lambda}x$.

(d) Sean λ_1 y λ_2 distintos eigenvalores de T con eigenvectores correspondientes x_1 y x_2. Entonces, utilizando el inciso (c), tenemos que

$$\begin{aligned}\lambda_1(x_1, x_2) &= (\lambda_1 x_1, x_2) = (T(x_1), x_2)\\ &= (x_1, T^*(x_2)) = (x_1, \bar{\lambda}_2 x_2) = \lambda_2(x_1, x_2).\end{aligned}$$

Como $\lambda_1 \neq \lambda_2$ concluimos que $(x_1, x_2) = 0$. ■

Corolario 1. *Sea* T *un operador normal en un espacio con producto interior* V, *y sea* β *una base ortonormal para* V. *Entonces* β *está formada por eigenvectores de* T *si y sólo si* β *está formada por eigenvectores de* T^*.

Como se mencionó anteriormente, existe un fuerte paralelismo entre el complejo conjugado de un número complejo y el adjunto de un operador lineal. (Véase, por ejemplo, el Teorema 7.10.) Los números reales pueden caracterizarse como aquellos números complejos que son iguales a sus complejos conjugados. Si consideramos la condición $T = T^*$ para un operador lineal, veremos que se tienen muchas de las propiedades de los números reales para dichos operadores. De hecho, veremos (en el Ejercicio 5) que todo operador puede escribirse en la forma $T_1 + iT_2$, donde T_1 y T_2 satisfacen la condición anterior. Asimismo, todos estos operadores tienen únicamente eigenvalores reales.

Definiciones. *Sea* V *un espacio con producto interior, y sea* T *un operador lineal en* V. T *se denomina operador* autoadjunto *(o* Hermitiano*) si* $T = T^*$. *Una matriz* A *de* $n \times n$ *es* autoadjunta *(o* Hermitiana*) si* $A = A^*$.

Así, para el caso de matrices reales, ser autoadjunta equivale a ser simétrica.

Es fácil ver que si β es una base ortonormal finita para V, entonces T es autoadjunto si y sólo si $[T]_\beta$ es autoadjunta.

Nótese también que cualquier matriz diagonal que tenga al menos un elemento no real es normal pero no autoadjunta. Por supuesto, las matrices autoadjuntas son normales.

Ejemplo 20. Sea $V = R^2$; entonces $V = W_1 \oplus W_2$, donde $W_1 = L(\{(1, 1)\})$ y $W_2 = L(\{(0, 1)\})$. Sea T la proyección en W_1 tal que $N(T) = W_2$; esto es, $T(a, b) = (a, a)$. Si $\beta = \{e_1, e_2\}$, entonces

$$A = [T]_\beta = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

Dado que

$$A^* = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix},$$

un cálculo sencillo muestra que $AA^* \neq A^*A$. Así pues, T no es ni autoadjunto ni normal.

Veremos en la Sección 7.5 que un tipo especial de proyección T, llamada "proyección ortogonal" (la que tiene la propiedad de que $R(T) = N(T)^{\perp}$), es siempre autoadjunta.

Corolario 2. *Sea* T *un operador lineal autoadjunto en un espacio con producto interior* V. *Si* λ *es un eigenvalor de* T, *entonces* λ *es un número real.*

DEMOSTRACIÓN. Sea x un eigenvector correspondiente al eigenvalor λ. Por el inciso (c) del teorema anterior tenemos que

$$\lambda x = T(x) = T^*(x) = \bar{\lambda} x.$$

Como $x \neq 0$, tenemos que $\lambda = \bar{\lambda}$; por lo tanto λ es real. ■

En el Teorema 7.16 demostraremos que una cierta clase de operador lineal posee siempre un eigenvalor.

El lector observará que las hipótesis de este teorema están divididas en dos casos, $F = R$ y $F = C$. La razón de esto se hará evidente en el punto de la demostración donde deseemos obtener un cero del polinomio característico, ya que, aunque muchos polinomios de valor real no tienen ceros (reales), el teorema fundamental del álgebra (Apéndice D) garantiza este resultado para los polinomios de valor complejo.

Teorema 7.16. *Sea* V *un espacio vectorial dimensionalmente finito sobre* F, *y sea* T *un operador lineal en* V.

(a) *Si* V *es un espacio vectorial complejo (esto es, si* $F = C$*), entonces* T *tiene un eigenvalor.*

(b) *Si* V *es un espacio real con producto interior (o sea, si* $F = R$*) y* T *es autoadjunto, entonces* T *tiene un eigenvalor (real).*

DEMOSTRACIÓN. Supóngase que $\dim(V) = n$, y sea f el polinomio característico de T.

(a) Si $F = C$, entonces el teorema fundamental del álgebra garantiza que f tiene un cero. Por tanto, T tiene un eigenvalor.

(b) Si V es un espacio real con producto interior, sea β una base ortonormal para V. Entonces $A = [T]_\beta$ es autoadjunta y tiene elementos reales.

Defínase a $T_A\colon C^n \to C^n$ mediante $T_A(x) = Ax$. De (a) se tiene que T_A tiene un eigenvalor λ. Como la matriz de T_A en la base ordenada están-

dar para C^n es A, tenemos que T_A es autoadjunto y, por lo tanto, según el Corolario 2 del Teorema 7.15, λ es real. Luego entonces, el polinomio $f(t) = \det(A - tI)$ tiene el cero real λ y entonces T tiene al eigenvalor λ. ■

Teorema 7.17(C). *Sea V un espacio con producto interior, complejo y dimensionalmente finito, y sea T un operador lineal en V. Entonces T es normal si y sólo si V tiene una base ortonormal formada por eigenvectores de T.*

Teorema 7.17(R). *Sea V un espacio con producto interior, real y dimensionalmente finito, y sea T un operador lineal en V. Entonces T es autoadjunto si y sólo si V tiene una base ortonormal formada por eigenvectores de T.*

DEMOSTRACIÓN. Supondremos primero que T es o normal o autoadjunto y luego obtendremos la base ortonormal adecuada. La demostración se hará por inducción sobre $n = \dim(V)$.

Si $n = 1$, entonces $V = L(\{x\})$ para alguna $x \neq 0$. En este caso es obvio que $\{(1/\|x\|)x\}$ es una base ortonormal formada por un eigenvector de T.

Ahora supóngase que el resultado es cierto para operadores normales [autoadjuntos] en espacios con producto interior de dimensión $n - 1$. Demostraremos que el resultado es cierto para el operador T en V.

Por el Teorema 7.16, T tiene un eigenvalor λ_1; sea x_1 un eigenvector asociado. Supondremos que $\|x_1\| = 1$. Sea $W = L(\{x_1\})$. De acuerdo con el Teorema 7.15, x_1 es también un eigenvector de T*, de manera que evidentemente W es T- y T*-invariante. Por el Ejercicio 6, $W^\perp$ es también T- y T*-invariante; por tanto, por el mismo ejercicio, $T_{W^\perp}$ es normal [autoadjunto] puesto que T lo es. Del Corolario 2 del Teorema 7.6 tenemos que $\dim(W^\perp) = n - 1$. Por lo tanto, podemos aplicar la hipótesis de inducción a $T_{W^\perp}$ para producir una base ortonormal $\{x_2, \ldots, x_n\}$ para $W^\perp$ formada por eigenvectores de $T_{W^\perp}$ y, por lo tanto, de T. Se infiere fácilmente $\{x_1, \ldots, x_n\}$ es la base ortonormal para V deseada.

La primera parte de la demostración, que es la más difícil, queda terminada. Ahora supongamos que $\{x_1, \ldots, x_n\}$ es una base ortonormal formada por eigenvectores de T con $T(x_i) = \lambda_i x_i$ para $1 \leq i \leq n$.

Si V es un espacio complejo con producto interior, entonces por el Teorema 7.15

$$(TT^*)(x_i) = T(\bar{\lambda}_i x_i) = \bar{\lambda}_i T(x_i) = \bar{\lambda}_i \lambda_i x_i = |\lambda_i|^2 x_i \qquad \text{para } 1 \leq i \leq n.$$

Análogamente,

$$(T^*T)(x_i) = |\lambda_i|^2 x_i \qquad \text{para } 1 \leq i \leq n.$$

Por lo tanto, T es normal.

Por otra parte, si V es un espacio real con producto interior, entonces λ_i es real para $1 \leq i \leq n$. Así,

$$T(x_i) = \lambda_i x_i = \bar{\lambda}_i x_i = T^*(x_i) \qquad \text{para } 1 \leq i \leq n$$

y, por lo tanto, T es autoadjunto. ■

Con el objeto de ver por qué la condición de normalidad en un espacio real con producto interior no es suficiente para garantizar ni siquiera un eigenvector, basta con que consideremos las rotaciones. Sea $0 < \theta < \pi$ y defínase a T: $R^2 \to R^2$ como la rotación en un ángulo θ. La matriz de T en la base ordenada estándar es

$$A = \begin{pmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{pmatrix}.$$

Es fácil ver que $AA^* = I = A^*A$ pero que $A \neq A^*$. Es geométricamente evidente que tal rotación no tiene eigenvectores.

Concluiremos esta sección con un ejemplo de un operador normal en un espacio complejo con producto interior que ¡no tiene eigenvectores! Así pues, la hipótesis de que V sea dimensionalmente finito es crucial en el Teorema 7.17(C), así como en el Teorema 7.16.

Ejemplo 21. Considérese el espacio con producto interior H anteriormente definido, y sea $x_k = e^{ikr}$. Supóngase que $V = L(\{x_k\colon k \text{ es un entero}\})$. Evidentemente $\beta = \{x_k\colon k \text{ es un entero}\}$ es una base ortonormal de V. Selecciónense ahora operadores lineales T y U en V tales que $T(x_k) = x_{k+1}$ y $U(x_k) = x_{k-1}$ para todo entero k. Entonces

$$\begin{aligned}(T(x_i), x_j) &= (x_{i+1}, x_j) = \delta_{(i+1)j} = \delta_{i(j-1)} \\ &= (x_i, x_{j-1}) = (x_i, U(x_j)).\end{aligned}$$

Se infiere que $U = T^*$. Además, $TT^* = I = T^*T$ y entonces T es normal Para cualquier elemento $x \in V$ tenemos que

$$x = \sum_{i=-k}^{k} a_i x_i$$

para alguna k y escalares a_i, y entonces

$$T(x) = \sum_{i=-k}^{k} a_i x_{i+1}.$$

Como β es independiente, se infiere que T no tiene eigenvectores.

La condición de que $TT^* = I = T^*T$, la cual apareció en los últimos dos ejemplos, será considerada en la sección siguiente.

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas. Supóngase que los espacios con producto interior subyacentes son dimensionalmente finitos.

   (a) Todo operador autoadjunto es normal.
   (b) Los operadores tienen los mismos eigenvectores que sus adjuntos.
   (c) Si T es un operador en un espacio con producto interior V, entonces T es normal si y sólo si $[T]_\beta$ es normal, donde β es cualquier base ordenada para V.
   (d) Una matriz A es normal si y sólo si L_A es normal.
   (e) Los eigenvalores de un operador autoadjunto deben ser todos reales.
   (f) Los operadores identidad y nulo son autoadjuntos.
   (g) Todo operador normal es diagonalizable.
   (h) Todo operador autoadjunto es diagonalizable.

2. Para cada uno de los siguientes operadores lineales, determinar si son normales, autoadjuntos o ninguno de los dos.

   (a) T: $R^2 \to R^2$ definido mediante $T(a, b) = (2a - 2b, -2a + 5b)$
   (b) T: $C^2 \to C^2$ definido mediante $T(a, b) = (2a + ib, a + 2b)$
   (c) T: $P_2(R) \to P_2(R)$ definido mediante $T(f) = f'$.

   Para el inciso (a), encontrar una base ortonormal para R^2 formada por eigenvectores de T.

3. Sean T y U operadores autoadjuntos en un espacio con producto interior. Demostrar que TU es autoadjunto si y sólo si TU = UT.

4. Demostrar el inciso (b) del Teorema 7.15.

5. Sea V un espacio complejo con producto interior, y sea T un operador lineal en V. Defínase

$$T_1 = \frac{1}{2}(T + T^*) \quad y \quad T_2 = \frac{1}{2i}(T - T^*).$$

   (a) Demostrar que T_1 y T_2 son autoadjuntos y que $T = T_1 + iT_2$.
   (b) Suponer también que $T = U_1 + iU_2$, donde U_1 y U_2 son autoadjuntos. Demostrar que $U_1 = T_1$ y $U_2 = T_2$.
   (c) Demostrar que T es normal si y sólo si $T_1T_2 = T_2T_1$.

6. Sea T un operador lineal en un espacio con producto interior V, y sea W un subespacio T-invariante de V. Demostrar las cuestiones siguientes.

   (a) Si T es autoadjunto, entonces también T_W lo es.
   (b) $W^\perp$ es T*-invariante.

(c) Si W es T- y T*-invariante, entonces $(T_W)^* = (T^*)_W$.
(d) Si W es T- y T*-invariante y T es normal, entonces T_W es normal.

7. Sea T un operador normal en un espacio complejo con producto interior dimensionalmente finito V, y sea W un subespacio de V. Demostrar que si W es T-invariante, entonces W es también T*-invariante. *Sugerencia:* Utilizar el Ejercicio 10(d) de la Sección 5.4.

8. Sea T un operador normal en un espacio con producto interior dimensionalmente finito V. Demostrar que $N(T) = N(T^*)$ y $R(T) = R(T^*)$. *Sugerencia:* Utilizar el Teorema 7.15 y el Ejercicio 12 de la Sección 7.3.

9. Sea T un operador autoadjunto en un espacio con producto interior dimensionalmente finito V. Demostrar que para toda $x \in V$

$$||T(x) \pm ix||^2 = ||T(x)||^2 + ||x||^2.$$

Deducir que $(T - iI)$ es invertible y que $[(T - iI)^{-1}]^* = (T + iI)^{-1}$.

10. Supóngase que T es un operador lineal en un espacio complejo con producto interior (no necesariamente dimensionalmente finito) V con un adjunto T*. Demostrar que

(a) Si T es autoadjunto, entonces $(T(x), x)$ es real para toda $x \in V$.
(b) Si T satisface a $(T(x), x) = 0$ para toda $x \in V$, entonces $T = T_0$. *Sugerencia:* Sustituir a x por $x + y$ y luego por $x + iy$ y expándanse los productos interiores resultantes.
(c) Si $(T(x), x)$ es real para toda $x \in V$, entonces $T = T^*$.

11. Sea A una matriz real de $n \times n$. Se dice que A es una matriz *Gramiana* si existe una matriz real B (cuadrada) tal que $A = B^tB$. Demostrar que A es una matriz Gramiana si y sólo si A es simétrica y todos sus eigenvalores son no negativos. *Sugerencia:* Aplicar el Teorema 7.17(R) a L_A para obtener una base ortonormal $\{x_1, \ldots, x_n\}$ de eigenvectores con los eigenvalores asociados $\lambda_1, \ldots, \lambda_n$.

12. Sea T un operador autoadjunto en un espacio con producto interior n-dimensional V, y sea $A = [T]_\beta$, donde β es una base ortonormal para V. Se dice que T es *definido* [*semidefinido*] *positivo* si $(T(x), x) > 0$ para toda $x \neq 0$ [$(T(x), x) \geq 0$ para toda x]. Demostrar

(a) T es definido [semidefinido] positivo si y sólo si todos sus eigenvalores son positivos [no negativos].
(b) T es definido [semidefinido] positivo si y sólo si L_A también lo es.
(c) T es definido positivo si y sólo si

$$\sum_{i,j} A_{ij}a_i\overline{a_j} > 0 \qquad \text{para todas las } n\text{-adas no nulas } (a_1, \ldots, a_n).$$

(d) T es semidefinido positivo si y sólo si A es una matriz Gramiana (tal como se definió en el Ejercicio 11).

La composición de dos operadores definidos positivos ¿es definido positivo?

13. *Diagonalización simultánea.*

(a) Sea V un espacio con producto interior, real y dimensionalmente finito, y sean U y T operadores autoadjuntos en V tales que $UT = TU$. Demostrar que existe una base ortonormal para V formada por vectores que son eigenvectores de U y de T. (La versión compleja de este resultado aparece como el Ejercicio 10 de la Sección 7.9.) *Sugerencia:* Para cualquier eigenespacio $W = E_\lambda$ de T tenemos que W es T- y U-invariante. Por el Ejercicio 6 tenemos que $W^\perp$ es T- y U-invariante. Aplicar el Teorema 7.17(R) y el Teorema 7.6.

(b) Enunciar y demostrar los resultados análogos acerca de matrices simétricas (reales) conmutativas.

14. Sea T un operador lineal en un espacio con producto interior dimensionalmente finito V. Demostrar lo siguiente.

(a) Si $\beta = \{x_1, \ldots, x_n\}$ es una base ordenada para V, entonces $[T]_\beta$ es triangular superior si y sólo si $T(x_j) \in L(\{x_1, \ldots, x_j\})$ para $j = 1, \ldots, n$.

(b) Si V es un espacio complejo con producto interior, entonces existe una base ortonormal γ para V tal que $[T]_\gamma$ es triangular superior. *Sugerencia:* Utilizar inducción sobre $n = \dim(V)$. Escójase un eigenvector x de T^* y sea $W = L(\{x\})$. Aplíquese la hipótesis de inducción a $W^\perp$ que, de acuerdo con el Ejercicio 6(b), es T-invariante.

(c) Toda matriz compleja A es similar a una matriz triangular superior.

15. Demostrar el *teorema de Cayley-Hamilton* para una matriz compleja A de $n \times n$; esto es, si f es el polinomio característico de A, demostrar que $f(A) = 0$. *Sugerencia:* De acuerdo con el inciso (c) del Ejercicio 14, demostrar que es posible suponer que A es triangular superior, en cuyo caso

$$f(t) = \prod_{i=1}^{n} (A_{ii} - t).$$

Ahora bien, si $T = L_A$, tenemos que $(A_{jj}I - T)(x_j) \in L(\{x_1, \ldots, x_{j-1}\})$, donde $\{x_1, \ldots, x_n\}$ es la base ordenada estándar para C^n.

7.6* EL CONDICIONAMIENTO Y EL COCIENTE DE RAYLEIGH

En la Sección 3.4 estudiamos técnicas específicas que nos permitieron resolver sistemas de ecuaciones lineales de la forma $AX = b$ donde A es una matriz de $m \times n$ y b es un vector de $m \times 1$. Tales sistemas surgen

a menudo en las aplicaciones a la vida real. Los coeficientes del sistema se obtienen a partir de datos experimentales y en muchos casos m y n son tan grandes que es necesario utilizar una computadora para obtener la solución. Así pues, debemos considerar dos tipos de errores. Primero, hay errores experimentales que surgen en la recopilación de la información puesto que ningún instrumento puede permitir realizar medidas totalmente exactas. Segundo, las computadoras introducirán errores de redondeo. Intuitivamente se puede sentir que cambios relativamente pequeños en los coeficientes del sistema provocarán errores relativamente pequeños en la solución. Un sistema que tiene esta propiedad se denomina *bien condicionado;* de lo contrario el sistema se llama *pobremente condicionado.*

Consideraremos ahora algunos ejemplos de estos tipos de errores, concentrándonos principalmente en los cambios en b más bien que en los cambios de los elementos de A. Además, supondremos que A es cuadrada, compleja (o real) e invertible puesto que este es el caso que se encuentra con más frecuencia en las aplicaciones.

Ejemplo 22. Considérese el sistema

$$\begin{cases} x_1 + x_2 = 5 \\ x_1 - x_2 = 1. \end{cases}$$

La solución del sistema será

$$\begin{pmatrix} 3 \\ 2 \end{pmatrix}.$$

Ahora supóngase que cambiamos un poco al sistema y considérese el nuevo sistema

$$\begin{cases} x_1 + x_2 = 5 \\ x_1 - x_2 = 1.0001. \end{cases}$$

Este sistema modificado tiene la solución

$$\begin{pmatrix} 3.00005 \\ 1.99995 \end{pmatrix}.$$

Vemos que una modificación de 10^{-4} en uno de los coeficientes ha modificado en menos de 10^{-4} a cada una de las coordenadas de la nueva solución. Más generalmente, el sistema

$$\begin{cases} x_1 + x_2 = 5 \\ x_1 - x_2 = 1 + \delta \end{cases}$$

tiene como solución a

$$\begin{pmatrix} 3 + \dfrac{\delta}{2} \\ 2 - \dfrac{\delta}{2} \end{pmatrix}.$$

Por lo tanto, pequeños cambios en b introducen pequeños cambios en la solución. Por supuesto, estamos realmente interesados en los "cambios relativos" puesto que un cambio en la solución de, por ejemplo, 10 se considera grande si la solución original es del orden de 10^{-2} pero pequeño si la solución original es del orden de 10^6.

Introduciremos la notación δb para representar al vector $b' - b$, donde b es el vector del sistema original y b' es el vector del sistema modificado. Así, en el Ejemplo 22, tenemos que

$$\delta b = \begin{pmatrix} 5 \\ 1 + \delta \end{pmatrix} - \begin{pmatrix} 5 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \delta \end{pmatrix}.$$

Definiremos ahora el *cambio relativo* en b como el escalar $\|\delta b\|/\|b\|$, donde $\|\cdot\|$ denota la norma estándar en C^n (o R^n); esto es $\|b\| = \sqrt{(b, b)}$. La mayor parte de lo que sigue, sin embargo, es cierto para cualquier norma. Definiciones semejantes se cumplen para el *cambio relativo* en x. Así, en el Ejemplo 22,

$$\frac{\|\delta b\|}{\|b\|} = \frac{|\delta|}{\sqrt{26}}, \quad \text{y} \quad \frac{\|\delta x\|}{\|x\|} = \frac{\left\|\begin{pmatrix} 3 + (\delta/2) \\ 2 - (\delta/2) \end{pmatrix} - \begin{pmatrix} 3 \\ 2 \end{pmatrix}\right\|}{\left\|\begin{pmatrix} 3 \\ 2 \end{pmatrix}\right\|} = \frac{|\delta|}{\sqrt{26}}.$$

De manera que el cambio relativo en x es igual, coincidentemente, al cambio relativo en b y, por lo tanto, el sistema está bien condicionado.

Ejemplo 23. Considérese el sistema

$$\begin{cases} x_1 + x_2 = 3 \\ x_1 + 1.00001x_2 = 3.00001, \end{cases}$$

el cual tiene a

$$\begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

como solución. La solución para el sistema relacionado

$$\begin{cases} x_1 + x_2 = 3 \\ x_1 + 1.00001x_2 = 3.00001 + \delta \end{cases}$$

es

$$\begin{pmatrix} 2 - (10^5)\delta \\ 1 + (10^5)\delta \end{pmatrix}.$$

Por lo tanto,

$$\frac{\|\delta x\|}{\|x\|} = \left(\sqrt{\tfrac{2}{5}}\right) 10^5|\delta| \geq 10^4|\delta|,$$

mientras que

$$\frac{||\delta b||}{||b||} \approx \frac{|\delta|}{5}.$$

Tenemos que el cambio relativo en x es al menos ¡10^4 veces el cambio relativo en b! El sistema está muy pobremente condicionado. Obsérvese que las líneas rectas definidas por las dos ecuaciones de este sistema son casi coincidentes, de modo que una pequeña modificación en cualquiera de las rectas alteraría en gran medida el punto de intersección, esto es, la solución del sistema.

Para aplicar toda la fuerza de la teoría de las matrices autoadjuntas al estudio del condicionamiento, necesitamos tener la noción de norma de una matriz. (Ver Ejercicio 22 de la Sección 7.1 para otros resultados sobre normas.)

Definición. *Sea* A *una matriz compleja (o real) de* $n \times n$. *Defínase la* norma (Euclidiana) *de* A *mediante*

$$||A|| = \max_{x \neq 0} \frac{||Ax||}{||x||},$$

donde $x \in C^n$ *(o* R^n*).*

Vemos intuitivamente que $||A||$ representa la "ampliación máxima de un vector mediante la matriz A.

La cuestión si este máximo existe o no, así como el problema de cómo calcularlo, serán resueltos por medio del llamado "cociente de Rayleigh".

Definición. *Sea* B *una matriz autoadjunta de* $n \times n$. *El* cociente de Rayleigh *para* $x \neq 0$ *se define como el escalar* $R(x) = (Bx, x)/||x||^2$.

Teorema 7.18. *Para una matriz autoadjunta* B *tenemos que* $\max_{x \neq 0} R(x)$ *es el mayor eigenvalor de* B *y* $\min_{x \neq 0} R(x)$ *es el mínimo eigenvalor de* B.

DEMOSTRACIÓN. Por el Teorema 7.17 podemos seleccionar una base ortonormal $\{x_1, \ldots, x_n\}$ de eigenvectores de B tales que $Bx_i = \lambda_i x_i$, $1 \leq i \leq n$, donde $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. (Recuérdese que, por el Corolario 2 del Teorema 7.15, los eigenvalores de B son reales.) Ahora bien, para $x \in C^n$ (o R^n) existen escalares $a_1, \ldots, a_n$ tales que

$$x = \sum_{i=1}^{n} a_i x_i;$$

por lo tanto

$$R(x) = \frac{(Bx, x)}{\|x\|^2} = \frac{\left(\sum_{i=1}^{n} a_i\lambda_i x_i, \sum_{j=1}^{n} a_j x_j\right)}{\|x\|^2} = \frac{\sum_{i=1}^{n} \lambda_i |a_i|^2}{\|x\|^2} \leq \frac{\lambda_1 \sum_{i=1}^{n} |a_i|^2}{\|x\|^2} = \lambda_1.$$

Es fácil ver que $R(x_1) = \lambda_1$ por lo que, con esto, hemos demostrado la primera mitad del teorema. La segunda mitad se demuestra de una manera semejante. ■

Corolario 1. *Para cualquier matriz cuadrada* A, $\|A\|$ *es finita y, de hecho, es igual a* $\sqrt{\lambda}$, *donde* λ *es el mayor eigenvalor de* A*A.

DEMOSTRACIÓN. Sea B la matriz autoadjunta A^*A, y sea λ el mayor eigenvalor de B. Como, para $x \neq 0$,

$$0 \leq \frac{\|Ax\|^2}{\|x\|^2} = \frac{(Ax, Ax)}{\|x\|^2} = \frac{(A^*Ax, x)}{\|x\|^2} = \frac{(Bx, x)}{\|x\|^2} = R(x),$$

tenemos del Teorema 7.18 que $\|A\|^2 = \lambda$. ■

Obsérvese que la demostración del Corolario 1 muestra que todos los eigenvalores de A^*A son no negativos. Para el siguiente corolario necesitamos del lema siguiente.

Lema. *Para cualquier matriz cuadrada* A, λ *es un eigenvalor de* A*A *si y sólo si* λ *es un eigenvalor de* AA*.

DEMOSTRACIÓN. Sea λ un eigenvalor de A^*A. Si $\lambda = 0$, entonces A^*A no es invertible. Por lo tanto, A (y A^*) no es invertible, de manera que λ es también un eigenvalor de AA^*. La prueba de la recíproca es semejante.

Supóngase ahora que $\lambda \neq 0$. Entonces existe $x \neq 0$ tal que $A^*Ax = \lambda x$. Aplicando A a ambos lados tenemos que $(AA^*)(Ax) = \lambda(Ax)$. Como $Ax \neq 0$ (pues de lo contrario $\lambda x = 0$), tenemos que λ es un eigenvalor de AA^*. La recíproca se deja como ejercicio. ■

Corolario 2. *Sea* A *una matriz invertible. Entonces* $\|A^{-1}\| = 1/\sqrt{\lambda}$, *donde* λ *es el eigenvalor más pequeño de* A*A.

DEMOSTRACIÓN. Haremos uso de la observación de que λ es un eigenvalor de una matriz invertible si y sólo si λ^{-1} es un eigenvalor de su inversa.

Ahora bien, sean $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ los eigenvalores de A^*A, los cuales, por el lema, son los eigenvalores de AA^*. Entonces $\|A^{-1}\|^2$ es igual al mayor eigenvalor de $(A^{-1})^*A^{-1} = (AA^*)^{-1}$, el cual es igual a $1/\lambda_n$. ■

Para muchas aplicaciones, únicamente los eigenvalores mayor y menor son de interés. Por ejemplo, en el caso de los problemas de vibración, el eigenvalor más pequeño representa la mínima frecuencia a la que las vibraciones pueden ocurrir.

Veremos el papel de ambos eigenvalores en nuestro estudio de condicionamiento.

Ejemplo 24. Sea

$$A = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

Entonces

$$B = A^*A = \begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

Los eigenvalores de B son 3, 3 y 0. Por tanto, $||A|| = \sqrt{3}$. Para cualquier

$$x = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \neq 0,$$

podemos calcular $R(x)$ para la matriz B como

$$3 \geq R(x) = \frac{(Bx,\ x)}{||x||^2} = \frac{2(a^2 + b^2 + c^2 - ab + ac + bc)}{a^2 + b^2 + c^2}$$

para toda $a, b, c \in R$.

Ahora que sabemos que $||A||$ existe para toda matriz cuadrada, utilizaremos la desigualdad $||Ax|| \leq ||A|| \cdot ||x||$, la que se cumple para toda x.

Supóngase, para lo que sigue, que A es invertible, $b \neq 0$ y $Ax = b$. Para una δb dada, sea δx el vector que satisface a $A(x + \delta x) = b + \delta b$. Entonces $A(\delta x) = \delta b$, y así $\delta x = A^{-1}(\delta b)$. Por lo tanto,

$$||b|| = ||Ax|| \leq ||A|| \cdot ||x|| \quad \text{y} \quad ||\delta x|| = ||A^{-1}(\delta b)|| \leq ||A^{-1}|| \cdot ||\delta b||.$$

Así, tenemos que

$$\frac{||\delta x||}{||x||} \leq \frac{||A^{-1}|| \cdot ||\delta b|| \cdot ||A||}{||b||} = ||A|| \cdot ||A^{-1}|| \cdot \left(\frac{||\delta b||}{||b||}\right).$$

De una manera análoga,

$$\frac{1}{||A|| \cdot ||A^{-1}||}\left(\frac{||\delta b||}{||b||}\right) \leq \frac{||\delta x||}{||x||}.$$

El número $||A|| \cdot ||A^{-1}||$ se llama *número condicional* de A y se denota por *cond*(A). Debe hacerse notar que la definición de cond(A) depende de cómo definimos la norma de A. Existen muchas maneras razonables de

definir la norma de una matriz. De hecho, la única propiedad que utilizamos para establecer las desigualdades anteriores fue que $\|Ax\| \leq \|A\| \cdot \|x\|$ para toda x. Resumamos lo anterior con el teorema siguiente.

Teorema 7.19. *Para el sistema* $Ax = b$ *donde* A *es invertible y* $b \neq 0$, *tenemos los dos resultados siguientes:*

(a) $\dfrac{1}{cond(A)}\dfrac{\|\delta b\|}{\|b\|} \leq \dfrac{\|\delta x\|}{\|x\|} \leq cond(A)\dfrac{\|\delta b\|}{\|b\|}$ *(para cualquier norma* $\|\cdot\|$*).*

(b) $cond(A) = \sqrt{\dfrac{\lambda_1}{\lambda_n}}$, *donde* λ_1 *y* λ_2 *son los eigenvalores mayor y menor, respectivamente, de* A^*A. *(En este inciso suponemos que* $\|\cdot\|$ *es la norma Euclidiana, definida en esta sección.)*

DEMOSTRACIÓN. El enunciado (a) se obtiene de las desigualdades anteriores y el (b) de los Corolarios 1 y 2 del Teorema 7.18. ■

Del Teorema 7.19 es claro que cond$(A) \geq 1$. Se deja como ejercicio demostrar que cond$(A) = 1$ si y sólo si A es un múltiplo escalar de una matriz "unitaria" u "ortogonal", definida en la Sección 7.7. Además, puede demostrarse con algo de trabajo que en el inciso (a) se puede obtener la igualdad mediante una elección adecuada de b y δb.

Podemos darnos cuenta de inmediato del inciso (a) que si cond(A) se aproxima a 1, entonces estamos seguros que un error relativo pequeño en b obliga a un error relativo pequeño en x. Si cond(A) es grande, sin embargo, entonces el error relativo en x puede ser pequeño, aun cuando el error relativo en b sea grande, o bien el error relativo en x puede ser grande, ¡aun cuando el error relativo en b sea pequeño! En pocas palabras, cond(A) indica únicamente el potencial para errores relativos grandes.

Hasta ahora hemos considerado únicamente errores en el vector b. Si existe un error δA en la matriz de los coeficientes del sistema $AX = b$, la situación es más complicada. Por ejemplo, $A + \delta A$ puede dejar de ser invertible. Pero puede demostrarse bajo consideraciones adecuadas que se puede dar una cota para el error relativo en x en términos de cond(A). Por ejemplo, si $A + \delta A$ es invertible, Forsythe y Moler (Forsythe, George y Moler, Cleve B., *Computer Solution of Liner Algebraic Systems,* Prentice Hall, Inc., 1976, p. 23), demuestran que

$$\frac{\|\delta x\|}{\|x + \delta x\|} \leq \text{cond}(A)\,\frac{\|\delta A\|}{\|A\|}.$$

Debería mencionarse que, en la práctica, casi nunca se conoce cond(A), puesto que sería un gasto innecesario de tiempo calcular A^{-1} simplemente para determinar la norma. De hecho, si se utiliza una computadora para encontrar A^{-1}, la inversa de A así calculada únicamente se aproximará a A^{-1} y el error en la inversa calculada se verá afectado por la magnitud

de cond(A). ¡Y así caemos en un círculo vicioso! Existen, sin embargo, algunas situaciones en las cuales se puede encontrar una aproximación utilizable de cond(A). Así pues, en la mayor parte de los casos, la estimación del error relativo en x se basa en la estimación de cond(A).

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.

(a) Si $AX = b$ está bien condicionada, entonces cond(A) es pequeño.
(b) Si cond(A) es grande, entonces $AX = b$ está pobremente condicionada.
(c) Si cond(A) es pequeño, entonces $AX = b$ está bien condicionada.
(d) La norma de A es igual al cociente de Rayleigh.
(e) La norma de A es siempre igual al mayor eigenvalor de A.

2. Calcular las normas de las matrices siguientes.

(a) $\begin{pmatrix} 4 & 0 \\ 1 & 3 \end{pmatrix}$ (b) $\begin{pmatrix} 5 & 3 \\ -3 & 3 \end{pmatrix}$

(c) $\begin{pmatrix} 1 & \frac{-2}{\sqrt{3}} & 0 \\ 0 & \frac{-2}{\sqrt{3}} & 1 \\ 0 & \frac{2}{\sqrt{3}} & 1 \end{pmatrix}$

3. Demostrar que si B es simétrica, entonces $\|B\|$ es el mayor eigenvalor de B.

4. Sean A y A^{-1} las siguientes:

$$A = \begin{pmatrix} 6 & 13 & -17 \\ 13 & 29 & -38 \\ -17 & -38 & 50 \end{pmatrix} \quad \text{y} \quad A^{-1} = \begin{pmatrix} 6 & -4 & 1 \\ -4 & 11 & 7 \\ -1 & 7 & 5 \end{pmatrix}.$$

Los eigenvalores de A son aproximadamente 84.74, 0.2007 y 0.0588.

(a) Aproximar $\|A\|$, $\|A^{-1}\|$ y cond(A). (Obsérvese el Ejercicio 3 anterior.)
(b) Supóngase que tenemos vectores x y $\tilde{x}$ tales que $Ax = b$ y $\|b - A\tilde{x}\| \leq 0.001$. Utilizar (a) para determinar las cotas superiores para $\|\tilde{x} - A^{-1}b\|$ (el error absoluto) y $\|\tilde{x} - A^{-1}b\|/\|A^{-1}b\|$ (el error relativo).

5. Supóngase que x es la verdadera solución de $AX = b$ y que una computadora llega a una solución aproximada $\tilde{x}$. Si cond(A) = 100, $\|b\| = 1$ y $\|b - A\tilde{x}\| = 0.1$, obtener cotas superior e inferior para $\|x - \tilde{x}\|/\|x\|$.

6. Sea

$$B = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

Calcular

$$R\begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}, \qquad \|B\| \quad \text{y} \quad \text{cond}(B).$$

7. Sea B una matriz simétrica. Demostrar que $\min_{x \neq 0} R(x)$ es igual al eigenvalor más pequeño de B.

8. Demostrar que si λ es un eigenvalor de AA^*, entonces λ es un eigenvalor de A^*A. Esto completa la demostración del lema del Corolario 2 del Teorema 7.18.

9. Demostrar la desigualdad izquierda de (a) en el Teorema 7.19.

10. Demostrar que $\text{cond}(A) = 1$ si y sólo si A es un múltiplo escalar de una matriz unitaria u ortogonal, tal como ésta se define en la Sección 7.7.

11. (a) Sean A y B matrices cuadradas unitariamente equivalentes, tal como se definen en la Sección 7.7. Demostrar que $\|A\| = \|B\|$.

(b) Sea V un espacio con producto interior dimensionalmente finito, y sea T un operador lineal en V. Defínase

$$\|\mathsf{T}\| = \max_{x \neq 0} \frac{\|\mathsf{T}\|(x)\|}{\|x\|}.$$

Demostrar que $\|\mathsf{T}\| = \|[\mathsf{T}]_\beta\|$, donde β es cualquier base ortonormal de V.

(c) Sea V un espacio dimensionalmente finito con producto interior y con una base ortonormal $\{x_1, x_2, \ldots\}$. Sea T el operador lineal en V tal que $\mathsf{T}(x_k) = kx_k$. Demostrar que $\|\mathsf{T}\|$ (definido en (b)) no existe.

7.7 OPERADORES UNITARIOS Y ORTOGONALES Y SUS MATRICES

En esta sección continuaremos nuestra analogía entre los números complejos y los operadores lineales. Recordemos que el complejo conjugado de un número complejo actúa de una manera semejante al adjunto de un operador lineal. (Véase, por ejemplo, el Teorema 7.10.) Un número complejo z tiene una longitud de 1 si $z\bar{z} = 1$. En esta sección estudiaremos

a aquellos operadores lineales T en un espacio vectorial V tales que $TT^* = T^*T = I$. Veremos que éstos son justamente los operadores lineales que "preservan la longitud" en el sentido de que $\|T(x)\| = \|x\|$ para toda $x \in V$. También demostraremos que, en un espacio complejo con producto interior, dimensionalmente finito, éstos son los operadores normales cuyos eigenvalores tienen todos un valor absoluto de 1.

En capítulos anteriores estuvimos interesados en estudiar aquellas funciones que conservan la estructura del espacio subyacente. En particular, los operadores lineales conservan las operaciones de suma vectorial y de multiplicación por escalares y los isomorfismos conservan toda la estructura del espacio vectorial. Es ahora normal considerar aquellos operadores lineales T en un espacio con producto interior que conservan la longitud; es decir, $\|T(x)\| = \|x\|$ para toda x. Veremos que, de hecho, esta condición garantiza que T preserva el producto interior.

Definiciones. *Sea* V *un espacio con producto interior (sobre* F*), y sea* T *un operador lineal en* V*. Si* $\|T(x)\| = \|x\|$ *para toda* $x \in V$*, llamamos a* T *un* operador unitario *si* $F = C$ *y un* operador ortogonal *si* $F = R$.

Evidentemente, cualquier rotación o reflexión en R^2 preserva la longitud y, por lo tanto, es un operador ortogonal. Estudiaremos en la próxima sección a estos operadores con mucho más detalle.

Ejemplo 25. Sea $V = H$ y sea $h \in V$ con $|h(x)| = 1$ para toda x. Defínase a $T: V \to V$ mediante $T(f) = hf$. Entonces

$$\|T(f)\|^2 = \|hf\|^2 = \frac{1}{2\pi}\int_0^{2\pi} h(t)f(t)\overline{h(t)f(t)}\,dt = \|f\|^2$$

por el hecho de que $|h(t)|^2 = 1$ para toda t. Por tanto, T es un operador unitario.

Teorema 7.20. *Sea* V *un espacio con producto interior, dimensionalmente finito, y sea* T *un operador lineal en* V*. Entonces son equivalentes las siguientes condiciones.*

(a) $TT^* = T^*T = I$.
(b) $(T(x), T(y)) = (x, y)$ *para toda* $x, y \in V$.
(c) *Si* β *es una base ortonormal para* V*, entonces* $T(\beta)$ *es una base ortonormal para* V.
(d) *Existe una base ortonormal* β *para* V *tal que* $T(\beta)$ *es una base ortonormal para* V.
(e) $\|T(x)\| = \|x\|$ *para toda* $x \in V$.

Por lo tanto, todas las condiciones anteriores son equivalentes a la definición de un operador unitario u ortogonal. De (a) se deduce que todo operador unitario u ortogonal es normal.

Antes de probar el teorema, primero demostraremos el siguiente lema. Compárese éste con el Ejercicio 10(b) de la Sección 7.5.

Lema. *Sea* V *un espacio con producto interior, dimensionalmente finito, y sea* U *un operador autoadjunto en* V. *Si* $(x, U(x)) = 0$ *para toda* $x \in V$, *entonces* $U = T_0$.

DEMOSTRACIÓN. Por el Teorema 7.17 podemos escoger una base ortonormal β de eigenvectores de U. Si $x \in \beta$ entonces $U(x) = \lambda x$ para alguna λ. Entonces

$$0 = (x, U(x)) = (x, \lambda x) = \bar{\lambda}(x, x),$$

y $\bar{\lambda} = 0$. Por lo tanto, $U(x) = 0$ para toda $x \in \beta$ y, finalmente, $U = T_0$. ■

DEMOSTRACIÓN DEL TEOREMA 7.20. Primero demostraremos que (a) implica a (b).

Sea $x, y \in V$. Entonces $(x, y) = ((T^*T)(x), y) = (T(x), T(y))$.

En segundo lugar, demostraremos que (b) implica a (c). Sea $\beta = \{x_1, \ldots, x_n\}$ una base ortonormal para V. Entonces $T(\beta) = \{T(x_1), \ldots, T(x_n)\}$. Ahora bien, $(T(x_i), T(x_j)) = (x_i, x_j) = \delta_{ij}$. Así, $T(\beta)$ es una base ortonormal de V.

El que (c) implica a (d) es evidente.

Ahora demostraremos que (d) implica a (e). Sea $x \in V$ y sea $\beta = \{x_1, \ldots, x_n\}$. Tenemos que

$$x = \sum_{i=1}^{n} a_i x_i$$

para algunos escalares a_i y entonces, como β es ortonormal,

$$\|x\|^2 = \left(\sum_{i=1}^{n} a_i x_i, \sum_{j=1}^{n} a_j x_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \overline{a_j}(x_i, x_j) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \overline{a_j} \delta_{ij} = \sum_{i=1}^{n} |a_i|^2$$

Haciendo las mismas operaciones a

$$T(x) = \sum_{i=1}^{n} a_i T(x_i)$$

y utilizando el hecho de que $T(\beta)$ también es ortonormal, obtenemos

$$\|T(x)\|^2 = \sum_{i=1}^{n} |a_i|^2.$$

Por tanto $\|T(x)\| = \|x\|$.

Finalmente, demostraremos que (e) implica a (a). Para cualquier $x \in V$ tenemos que

$$(x, x) = \|x\|^2 = \|T(x)\|^2 = (T(x), T(x)) = (x, (T^*T)(x)).$$

Así, $(x, (\text{I} - \text{T}^*\text{T})(x)) = 0$ para toda $x \in \text{V}$. Sea $\text{U} = \text{I} - \text{T}^*\text{T}$; entonces U es autoadjunta y $(x, \text{U}(x)) = 0$ para toda $x \in \text{V}$. Luego, de acuerdo con el lema, tenemos que $\text{T}_0 = \text{U} = \text{I} - \text{T}^*\text{T}$ y por lo tanto $\text{T}^*\text{T} = \text{I}$. Así pues, como V es dimensionalmente finito, $\text{T}^* = \text{T}^{-1}$ y entonces $\text{TT}^* = \text{I}$. ■

De la definición se infiere directamente que el valor absoluto de todo eigenvalor de un operador unitario u ortogonal es 1. De hecho, algo más es cierto.

Corolario 1. *Sea T un operador lineal en V un espacio real dimensionalmente finito con producto interior. V tiene una base ortonormal de eigenvectores de T con eigenvalores correspondientes de valor absoluto 1 si y sólo si T es autoadjunto y ortogonal.*

DEMOSTRACIÓN. Supóngase que V tiene una base ortonormal $\{x_1, \ldots, x_n\}$ tal que $\text{T}(x_i) = \lambda_i x_i$ y $|\lambda_i| = 1$ para toda i. Por el Teorema 7.17(R), T es autoadjunto. Entonces $(\text{TT}^*)(x_i) = \text{T}(\lambda_i x_i) = \lambda_i \lambda_i x_i = \lambda_i^2 x_i = x_i$ para cada i, de modo que $\text{TT}^* = \text{I}$ y por el inciso (a) del Teorema 7.20, T es ortogonal.

Si T es autoadjunto, entonces por el Teorema 7.17(R) tenemos que V posee una base ortonormal $\{x_1, \ldots, x_n\}$ tal que $\text{T}(x_i) = \lambda_i x_i$ para toda i. Como T es ortogonal, tenemos que $|\lambda_i| \cdot \|x_i\| = \|\lambda_i x_i\| = \|\text{T}(x_i)\| = \|x_i\|$, y entonces $|\lambda_i| = 1$ para cada i. ■

Corolario 2. *Sea T un operador lineal en V un espacio complejo dimensionalmente finito con producto interior. Entonces, V tiene una base ortonormal de eigenvectores de T con eigenvalores correspondientes cuyo valor absoluto es 1 si y sólo si T es unitario.*

DEMOSTRACIÓN. La demostración es semejante a la del Corolario 1. ■

Ejemplo 26. Sea $\text{T}: \text{R}^2 \to \text{R}^2$ una rotación por θ, donde $0 < \theta < \pi$. Es geométricamente evidente que T "preserva la longitud", esto es que $\|\text{T}(x)\| = \|x\|$ para toda $x \in \text{R}^2$. El hecho de que las rotaciones por un ángulo fijo conservan la perpendicularidad no sólo puede verse geométricamente, sino que ahora se infiere del inciso (b) del Teorema 7.20. Probablemente el hecho de que tal transformación preserve el producto interior no sea tan evidente geométricamente; sin embargo, este hecho lo obtenemos también a partir de (b). Finalmente, una inspección de la matriz

$$\begin{pmatrix} \cos\theta & -\text{sen}\,\theta \\ \text{sen}\,\theta & \cos\theta \end{pmatrix}$$

revela que T no es autoadjunto para la restricción dada para θ. Como ya lo mencionamos anteriormente, este hecho también se infiere de la observación geométrica de que T no tiene eigenvectores y del Teorema 7.16.

Puede verse fácilmente de la matriz anterior que T* es una rotación por $-\theta$.

Ahora, examinaremos a las matrices que representan transformaciones unitarias y ortogonales.

Definiciones. *Supóngase que* A *es una matriz de* n × n *que satisface a* AA* = = A*A = I. *Llamamos a* A matriz unitaria *si tiene elementos complejos, y la llamamos* matriz ortogonal *si tiene elementos reales.*

Nótese que la condición $AA^* = I$ es equivalente a afirmar que los renglones $A_1, \ldots, A_n$ de A forman un conjunto ortonormal en F^n para

$$\delta_{ij} = I_{ij} = (AA^*)_{ij} = \sum_{k=1}^{n} A_{ik}(A^*)_{kj} = \sum_{k=1}^{n} A_{ik}\overline{A_{jk}} = (A_i, A_j).$$

Puede hacerse una observación semejante acerca de las columnas de A y de la condición $A^*A = I$.

Se infiere también de la definición anterior que si V es un espacio con producto interior y T es un operador lineal en V, entonces T es unitario [ortogonal] si y sólo si $[T]_\beta$ es unitaria [ortogonal] para alguna base ortonormal β de V.

Ejemplo 27. La matriz

$$\begin{pmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{pmatrix}$$

es claramente ortogonal. Se puede ver fácilmente que los renglones de la matriz forman un conjunto ortonormal en R^2.

Sabemos que para una matriz normal compleja A [autoadjunta real] existe una base ortonormal β para F^n formada por eigenvectores de A. Por lo tanto, A es similar a una matriz diagonal D. Por el Teorema 5.1, la matriz Q, cuyas columnas son los vectores de β, es tal que $D = Q^{-1}AQ$. Pero como las columnas de Q son una base ortonormal para F^n se infiere que Q es unitaria [ortogonal]. En este caso decimos que A es *unitariamente equivalente* [*ortogonalmente equivalente*] a D. Se ve fácilmente (ver Ejercicio 17), que esta relación es una relación de equivalencia en $M_{n\times n}(C)$ $[M_{n\times n}(R)]$. Más generalmente, A y B son unitariamente equivalentes [ortogonalmente equivalentes], si y sólo si existe una matriz unitaria [ortogonal] P tal que $A = P^*BP$.

El párrafo anterior ya ha demostrado la mitad de cada uno de los dos siguientes teoremas.

Teorema 7.21(C). *Sea* A *una matriz compleja de* n × n. *Entonces* A *es normal si y sólo si* A *es unitariamente equivalente a una matriz diagonal.*

Teorema 7.21(R). *Sea* A *una matriz real de* n × n. *Entonces* A *es autoadjunta si y sólo si* A *es ortogonalmente equivalente a una matriz real diagonal.*

DEMOSTRACIÓN. Por las observaciones anteriores, necesitamos demostrar únicamente que si A es unitariamente [ortogonalmente] equivalente a una matriz diagonal, entonces A es normal [autoadjunta].

Supóngase que $A = P^*DP$, donde P es una matriz unitaria y D es una matriz diagonal. Entonces $AA^* = (P^*DP)(P^*DP)^* = (P^*DP)(P^*D^*P) = P^*DID^*P = P^*DD^*P$.

Análogamente $A^*A = P^*D^*DP$. Como D es una matriz diagonal, sin embargo, tenemos que $DD^* = D^*D$. Entonces $AA^* = A^*A$.

El resto de la demostración se deja al lector. ■

Ejemplo 28. Sea

$$A = \begin{pmatrix} 3 & -4 \\ -4 & -3 \end{pmatrix};$$

entonces $A = A^*$. Los eigenvalores de L_A son $\lambda_1 = 5$ y $\lambda_2 = -5$. Correspondientes a cada uno de estos eigenvalores están los eigenvectores $y_1 = (-2, 1)$ y $y_2 = (1, 2)$. Como se esperaba, y_1 y y_2 son ortogonales. Sean

$$x_1 = \frac{1}{\sqrt{5}}(-2, 1), \quad x_2 = \frac{1}{\sqrt{5}}(1, 2) \quad \text{y} \quad \beta = \{x_1, x_2\}.$$

Entonces, β es una base ortonormal de eigenvectores de L_A. Como en el párrafo que precede al Teorema 7.21, sean

$$P = \begin{matrix} \frac{-2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{matrix} \quad \text{y} \quad D = \begin{pmatrix} 5 & 0 \\ 0 & -5 \end{pmatrix}.$$

Un cálculo sencillo muestra que $P^*AP = D$.

Una aplicación (secciones cónicas)

Como una aplicación del Teorema 7.21, consideremos la ecuación cuadrática

$$ax^2 + 2bxy + cy^2 + dx + ey + f = 0. \tag{12}$$

Para elecciones especiales de los coeficientes de la ecuación (12), obtenemos las distintas secciones cónicas. Por ejemplo, si $a = c = 1$, $b = d = e = 0$ y $f = -1$, obtenemos la ecuación de la circunferencia $x^2 + y^2 = 1$ con centro en el origen. Las secciones cónicas restantes, llamadas elipse, parábola e hipérbola, se obtienen mediante otra selección de coeficientes. La ausencia del término xy permite graficar fácilmente estas cónicas mediante el método de completar el cuadrado. Por ejemplo, $x^2 + 2x + y^2 +$

$+ 4y + 2 = 0$ puede reescribirse como $(x + 1)^2 + (y + 2)^2 = 3$, que es la ecuación de una circunferencia con centro en $(-1, -2)$ en el sistema de coordenadas cartesiano (x, y) y con radio $\sqrt{3}$. Si consideramos la transformación de coordenadas $(x, y) \to (x', y')$, donde $x' = x + 1$ y $y' = y + 2$, entonces nuestra ecuación se simplifica a $(x')^2 + (y')^2 = 3$. Este tipo de transfórmación (llamada *translación)* nos permite eliminar los términos en x y en y.

Ahora, nos concentraremos únicamente en la eliminación del término en xy. Para hacer esto consideremos la expresión

$$ax^2 + 2bxy + cy^2, \tag{13}$$

la que se denomina *forma cuadrática asociada* de la ecuación (12). Las formas cuadráticas serán estudiadas con más detalle en la Sección 7.11.

Si hacemos

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \quad \text{y} \quad X = \begin{pmatrix} x \\ y \end{pmatrix},$$

entonces la ecuación (13) puede reescribirse como $X^tAX = (AX, X)$. Por ejemplo, $3x^2 + 4xy + 6y$ puede escribirse como

$$X^t \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix} X.$$

El hecho de que A sea autoadjunta es crucial en nuestra exposición, puesto que, en virtud del Teorema 7.21, podemos escoger una matriz ortogonal P y una matriz diagonal D con elementos reales en la diagonal λ_1 y λ_2, tales que $P^tAP = D$. Ahora definamos a

$$X' = \begin{pmatrix} x' \\ y' \end{pmatrix}$$

mediante $X' = P^tX$ o, de manera equivalente, mediante $PX' = PP^tX = X$. Entonces

$$X^tAX = (PX')^tA(PX') = X'^t(P^tAP)X' = X'^tDX' = \lambda_1(x')^2 + \lambda_2(y')^2.$$

Así, la transformación $(x, y) \to (x', y')$ nos permite eliminar al término en xy en la ecuación (13) y, por lo tanto, en la ecuación (12).

Además, como P es ortogonal, tenemos, de acuerdo con el Ejercicio 20(c), que $\det(P) = \pm 1$. Si $\det(P) = -1$, podemos reemplazar a P por $Q = PE$, donde

$$E = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Entonces Q es ortogonal, $\det(Q) = 1$ y

$$Q^tAQ = E^tP^tAPE = E^tDE = \begin{pmatrix} \lambda_2 & 0 \\ 0 & \lambda_1 \end{pmatrix}.$$

Por lo tanto, también podemos suponer que $\det(P) = 1$. De acuerdo con el Ejercicio 20, tenemos que P (o P^t) representa geométricamente una rotación.

En resumen, el término en xy de la ecuación (12) puede ser eliminado mediante una rotación de los ejes x e y a los nuevos ejes x' e y' dada por $X = PX'$, donde P es una matriz ortogonal y $\det(P) = 1$. Además, los coeficientes de $(x')^2$ y $(y')^2$ son los eigenvalores de

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

Este resultado es una nueva forma de enunciar el *teorema de ejes principales* para $\mathbb{R}^2$. Los argumentos anteriores, por supuesto, se extienden fácilmente a ecuaciones cuadráticas de n variables. Por ejemplo, para el caso en que $n = 3$, mediante una selección especial de los coeficientes obtenemos las superficies cuádricas —el cono elíptico, el elipsoide, el paraboloide hiperbólico, etc.

Como ejemplo, considérese la ecuación cuadrática $2x^2 - 4xy + 5y^2 - 36 = 0$, para la cual la forma cuadrática asociada es $2x^2 - 4xy + 5y^2$. Con la notación anterior

$$A = \begin{pmatrix} 2 & -2 \\ -2 & 5 \end{pmatrix},$$

de manera que los eigenvalores de A son 6 y 1 con eigenvectores asociados

$$\begin{pmatrix} 1 \\ -2 \end{pmatrix} \quad \text{y} \quad \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Como es de esperarse (del Teorema 7.15), estos vectores son ortogonales. La correspondiente base ortonormal de eigenvectores es

$$\left\{ \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{-2}{\sqrt{5}} \end{pmatrix}, \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{pmatrix} \right\}.$$

Por lo tanto, si

$$P = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{-2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix},$$

entonces $P^tAP = \begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}$. Bajo la transformación $X = PX'$, o bien

$$x = \frac{1}{\sqrt{5}}x' + \frac{2}{\sqrt{5}}y' \qquad y = \frac{-2}{\sqrt{5}}x' + \frac{1}{\sqrt{5}}y',$$

tenemos la nueva forma cuadrática $6(x')^2 + (y')^2$. Así, la ecuación original $2x^2 - 4xy + 5y^2 - 36 = 0$ puede escribirse en la forma $6(x')^2 + (y')^2 = 36$, en la que se puede ver fácilmente que se trata de la ecuación de una elipse. (Véase Fig. 7.4.)

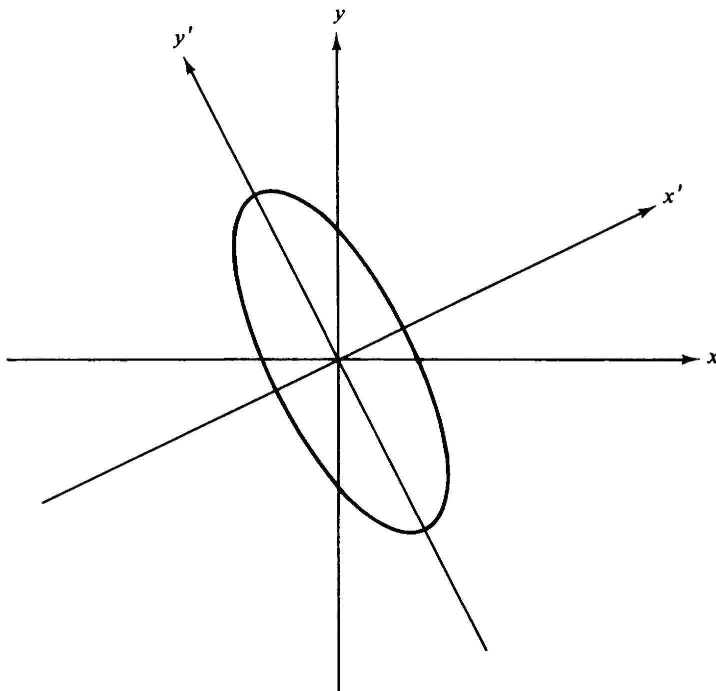


figura 7.4

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas. Supóngase que los espacios con producto interior subyacentes son dimensionalmente finitos.

   (a) Todo operador unitario es normal.
   (b) Todo operador ortogonal es diagonalizable.
   (c) Una matriz es unitaria si y sólo si es invertible.
   (d) Si dos matrices son unitariamente equivalentes también son semejantes.
   (e) La suma de dos matrices unitarias es unitaria.
   (f) El adjunto de un operador unitario es unitario.
   (g) Si T es un operador ortogonal en V, entonces $[T]_\beta$ es una matriz ortogonal para cualquier base ordenada β para V.

(h) Si todos los eigenvalores de un operador son 1, entonces el operador debe ser unitario u ortogonal.
(i) Un operador puede preservar la norma pero no el producto interior.

2. Para cada una de las matrices siguientes A, encontrar una matriz ortogonal o unitaria P y una matriz diagonal D tal que $P^*AP = D$.

(a) $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ (b) $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ (c) $A = \begin{pmatrix} 2 & 3-3i \\ 3+3i & 5 \end{pmatrix}$

(d) $A = \begin{pmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{pmatrix}$ (e) $A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$

3. Demostrar que el producto de operadores unitarios [ortogonales] es unitario [ortogonal].

4. Para $z \in C$ defínase a T_z: $C \to C$ mediante $\mathsf{T}_z(u) = zu$. Caracterizar aquellas z para las cuales T_z sea normal, autoadjunto o unitario.

5. ¿Cuáles de los siguientes pares de matrices son unitariamente equivalentes?

(a) $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ y $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ (b) $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ y $\begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}$

(c) $\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ y $\begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

(d) $\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ y $\begin{pmatrix} 1 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -i \end{pmatrix}$

6. Sea V el espacio con producto interior de las funciones continuas de valor complejo en [0, 1] con el producto interior

$$(f,\ g) = \int_0^1 f(t)\overline{g(t)}dt.$$

Sea $h \in \mathsf{V}$ y defínase a T: $\mathsf{V} \to \mathsf{V}$ mediante $\mathsf{T}(f) = hf$. Demostrar que T es un operador unitario si y sólo si $|h(t)| = 1$ para $0 \leq t \leq 1$.

7. Demostrar que si T es un operador unitario en un espacio con producto interior, dimensionalmente finito, entonces T tiene una "raíz cuadrada"; esto es, existe un operador unitario U tal que $\mathsf{T} = \mathsf{U}^2$.

8. Sea V un espacio con producto interior, y sea T: $\mathsf{V} \to \mathsf{V}$ un operador autoadjunto. Si $\mathsf{U} = (\mathsf{T} + i\mathsf{I})(\mathsf{T} - i\mathsf{I})^{-1}$, demostrar, utilizando el Ejercicio 9 de la Sección 7.5, que U es unitario.

9. Sea U un operador lineal en un espacio con producto interior dimensionalmente finito V. Si $\|U(x)\| = \|x\|$ para toda x en alguna base ortonormal para V, ¿debe U ser unitaria? Demostrar o dar un contra ejemplo.

10. Sea A una matriz compleja normal o real simétrica de $n \times n$ con eigenvalores $\lambda_1, \ldots, \lambda_n$ (no necesariamente distintos). Demostrar que

$$\operatorname{tr}(A) = \sum_{i=1}^{n} \lambda_i \quad \text{y} \quad \operatorname{tr}(A^*A) = \sum_{i=1}^{n} |\lambda_i|^2.$$

11. Encontrar una matriz ortogonal cuyo primer renglón sea $(\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$.

12. Sea A una matriz real simétrica o normal compleja de $n \times n$. Demostrar que

$$\det(A) = \prod_{i=1}^{n} \lambda_i,$$

donde los λ_i son los eigenvalores (no necesariamente distintos) de A.

13. Supóngase que A y B son matrices diagonalizables. Demostrar, afirmativa o negativamente, que A es semejante a B si y sólo si A y B son unitariamente equivalentes.

14. Sea U un operador unitario en un espacio con producto interior V, y sea W un subespacio U-invariante dimensionalmente finito de V. Demostrar que
    (a) $U(W) = W$.
    (b) $W^\perp$ es U-invariante.

    Contrastar a (b) con el resultado del Ejercicio 15.

15. Encontrar un ejemplo de un operador unitario U en un espacio con producto interior y un subespacio U-invariante W tal que $W^\perp$ no sea U-invariante.

16. Demostrar que una matriz que sea unitaria y triangular superior debe ser una matriz diagonal.

17. Demostrar que "es unitariamente equivalente a" es una relación de equivalencia en $M_{n\times n}(C)$.

18. Sea W un subespacio dimensionalmente finito de un espacio con producto interior V. Por el Teorema 7.6, $V = W \oplus W^\perp$. Defínase a $U\colon V \to V$ mediante $U(x_1 + x_2) = x_1 - x_2$, donde $x_1 \in W$ y $x_2 \in W^\perp$. Demostrar que U es un operador unitario autoadjunto.

19. Sea V un espacio con producto interior, dimensionalmente finito. Un operador lineal U en V, se llama *isometría parcial* si existe un subespacio W de V tal que $\|U(x)\| = \|x\|$ para toda $x \in W$ y $U(x) = 0$ para toda $x \in W^{\perp}$. Obsérvese que W *no* necesariamente tiene que ser U-invariante. Supóngase que U es tal operador y que $\{x_1, \ldots, x_k\}$ es una base ortonormal de W. Demostrar los incisos siguientes.

(a) $(U(x), U(y)) = (x, y)$ para toda $x, y \in W$. *Sugerencia:* Utilizar el Ejercicio 20 de la Sección 7.1.
(b) $\{U(x_1), \ldots, U(x_k)\}$ es una base ortonormal para R(U).
(c) Existe una base ortonormal γ para V tal que las primeras k columnas de $[U]_\gamma$ forman un conjunto ortonormal y las columnas restantes son nulas.
(d) Sea $\{y_1, \ldots, y_j\}$ una base ortonormal para $R(U)^{\perp}$. Sea $\beta = \{U(x_1), \ldots, U(x_k), y_1, \ldots, y_j\}$. Entonces β es una base ortonormal para V.
(e) Defínase a T como el operador lineal en V que satisface a $T(U(x_i)) = x_i (1 \leq i \leq k)$ y a $T(y_i) = 0 (1 \leq i \leq j)$. Demostrar que T está bien definido y que $T = U^*$. *Sugerencia:* Demostrar que $(U(x), y) = (x, T(y))$ para toda $x, y \in \beta$. Existen cuatro casos.
(f) Demostrar que U^* es una isometría parcial.

Este ejercicio continúa en el Ejercicio 9 de la Sección 7.9.

Una aplicación geométrica

La finalidad del ejercicio siguiente es emplear el conocimiento hasta ahora obtenido en este capítulo para caracterizar los llamados "movimientos rígidos" en R^2. Se puede pensar de una manera intuitiva en tal movimiento como una transformación que no afecta la forma de la figura bajo su acción; de ahí el nombre de "rígido". Por ejemplo, las reflexiones, las rotaciones y las translaciones $(x \rightarrow x + x_0)$ son ejemplos de movimientos rígidos. Veremos, de hecho, que todo movimiento rígido es una composición de estas tres transformaciones. La situación general en R^n será tratada en la Sección 7.8 y utilizará los resultados de este ejercicio.

20.* Sea V un espacio real con producto interior. Una función f: $V \rightarrow V$ se llama *movimiento rígido* si

$$\|f(x) - f(y)\| = \|x - y\| \qquad \text{para toda } x, y \in V.$$

Para tal función f, defínase a T: $V \rightarrow V$ mediante $T(x) = f(x) - f(0)$.

(a) Demostrar que T es lineal demostrando los cuatro incisos siguientes.

(i) $\|T(x)\| = \|x\|$ para toda $x \in V$.
(ii) $\|T(x) - T(y)\| = \|x - y\|$ para toda $x, y \in V$.

(iii) $(\mathsf{T}(x), \mathsf{T}(y)) = (x, y)$ para toda $x, y \in \mathsf{V}$. *Sugerencia:* Expándanse ambos lados de (ii) utilizando las propiedades de los productos interiores y luego iguálense los resultados.

(iv) $\|\mathsf{T}(x + ay) - \mathsf{T}(x) - a\mathsf{T}(y)\| = 0$ para toda $x, y \in \mathsf{V}$ y $a \in R$.

(b) Utilizar (a) para deducir que todo movimiento rígido es un operador ortogonal seguido de una translación.

(c) Demostrar que $\det(\mathsf{T}) = \pm 1$.

(d) Sea $\mathsf{V} = \mathsf{R}^2$ y sea β la base ordenada estándar para R^2. Demostrar que existe un ángulo $\theta (0 \leq \theta < 2\pi)$ tal que

$$[\mathsf{T}]_\beta = \begin{pmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{pmatrix} \qquad \text{si } \det(\mathsf{T}) = 1$$

y

$$[\mathsf{T}]_\beta = \begin{pmatrix} \cos\theta & \operatorname{sen}\theta \\ \operatorname{sen}\theta & -\cos\theta \end{pmatrix} \qquad \text{si } \det(\mathsf{T}) = -1.$$

Sugerencia: Utilizar el hecho de que las columnas de $[\mathsf{T}]_\beta$ forman un subconjunto ortonormal de R^2.

(e) Usar a (d) para deducir que todo movimiento rígido en R^2 es una rotación (con respecto al origen) seguida de una translación o una reflexión (con respecto al eje x) seguida de una rotación (con respecto al origen) seguida de una translación. *Sugerencia:* Obsérvese que

$$\begin{pmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} \cos\theta & \operatorname{sen}\theta \\ \operatorname{sen}\theta & -\cos\theta \end{pmatrix}.$$

21. Sean $y_1, \ldots, y_n$ vectores linealmente independientes en F^n y sean $x_1, \ldots, x_n$ vectores ortogonales obtenidos a partir de $y_1, \ldots, y_n$ mediante el proceso de ortogonalización de Gram-Schmidt. Sea $z_1, \ldots, z_n$ la base ortonormal obtenida al definir

$$z_k = \frac{x_k}{\|x_k\|}.$$

(a) Resolviendo la ecuación (1) de la Sección 7.2 para y_k en términos de z_k, demostrar que

$$y_k = \|x_k\| z_k + \sum_{j=1}^{k-1} (y_k, z_j) z_j \qquad (1 \leq k \leq n).$$

(b) Sean A y Q las matrices de $n \times n$ en las cuales las columnas k son y_k y z_k, respectivamente. Defínase a $R \in \mathsf{M}_{n\times n}(F)$ mediante

$$R_{jk} = \begin{cases} \|x_j\| & \text{si } j = k \\ (y_k, z_j) & \text{si } j < k \\ 0 & \text{si } j > k. \end{cases}$$

Demostrar que $A = QR$.

(c) Calcular Q y R como en el inciso (b) para la matriz de 3×3 cuyas columnas son los vectores y_1, y_2 y y_3, respectivamente, del Ejemplo 12 de la Sección 7.2.

(d) Como Q es unitaria [ortogonal] y R es triangular superior en el inciso (b), hemos demostrado que toda matriz invertible es el producto de una matriz unitaria [ortogonal] y una matriz triangular superior. Supóngase que $A \in \mathsf{M}_{n\times n}(F)$ es invertible y $A = Q_1R_1 = Q_2R_2$, donde Q_1, $Q_2 \in \mathsf{M}_{n\times n}(F)$ son unitarias y R_1, $R_2 \in \mathsf{M}_{n\times n}(F)$ son triangulares superiores. Demostrar que $D = R_2R_1^{-1}$ es una matriz unitaria diagonal. *Sugerencia:* Utilizar el Ejercicio 16.

(e) La descomposición descrita en el inciso (b) proporciona un método de ortogonalización para resolver un sistema lineal $AX = B$ donde A es invertible: Descomponer A en QR por el proceso de Gram-Schmidt (o por cualquier otro), donde Q sea unitaria y R sea triangular superior. Entonces $QRX = B$ y por lo tanto $RX = Q^*B$. Este último sistema puede resolverse fácilmente, puesto que R es triangular superior.

En un tiempo, a causa de su gran estabilidad, este método para resolver grandes sistemas de ecuaciones lineales por medio de una computadora se consideró como un método superior al de eliminación de Gauss, aun cuando requiere de tres veces más trabajo. (Posteriormente, sin embargo, J. H. Wilkinson demostró que si el método de eliminación de Gauss se lleva a cabo adecuadamente, entonces es casi tan estable como el método de ortogonalización.)

Emplear el método de ortogonalización y el inciso (c) para resolver el sistema

$$\begin{cases} x_1 + 2x_2 + 2x_3 = 1 \\ x_1 + 2x_3 = 11 \\ x_2 + x_3 = -1. \end{cases}$$

22. Encontrar nuevas coordenadas x', y' de manera que las siguientes formas cuadráticas puedan escribirse como $\lambda_1(x')^2 + \lambda_2(y')^2$.

(a) $x^2 + 4xy + y^2$
(b) $2x^2 + 2xy + 2y^2$

23. Considérese la expresión X^tAX, donde $X^t = (x, y, z)$ y A es como se definió en el Ejercicio 2(e). Encontrar un cambio de coordenadas x', y', z' de manera que la expresión anterior pueda escribirse en la forma $\lambda_1(x')^2 + \lambda_2(y')^2 + \lambda_3(z')^2$.

7.8* LA GEOMETRIA DE LOS OPERADORES ORTOGONALES

El Ejercicio 20 de la Sección 7.7 establece que cualquier movimiento rígido en un espacio real con producto interior es la composición de un operador

ortogonal seguido de una translación. Luego, para comprender detalladamente la geometría de los movimientos rígidos, es necesario analizar la estructura de los operadores ortogonales. Tal es la finalidad de esta sección. Como lo descubriremos, un operador ortogonal en un espacio real con producto interior dimensionalmente finito es el resultado de la composición de rotaciones y reflexiones. Principiaremos nuestra investigación con las definiciones de estos términos.

Definiciones. *Sea* T *un operador lineal en un espacio real con producto interior dimensionalmente finito* V. *El operador* T *se llama* rotación *si* T *es la identidad en* V *o si existe un subespacio bidimensional* W *de* V, *una base ortonormal* $\beta = \{x_1, x_2\}$ *para* W, *y un número real* θ *tal que*

$$T(x_1) = x_1 \cos\theta + x_2 \operatorname{sen}\theta, \qquad T(x_2) = -x_1 \operatorname{sen}\theta + x_2 \cos\theta,$$

y $T(y) = y$ *para toda* $y \in W^\perp$. *Dentro de este contexto* T *se denomina* rotación de W alrededor de $W^\perp$. *El subespacio* $W^\perp$ *se llama* eje de rotación.

En la Sección 2.1 se definieron las rotaciones para el caso especial donde $V = R^2$.

Definiciones. *Sea* T *un operador lineal en un espacio real con producto interior dimensionalmente finito* V. *El operador* T *se llama* reflexión *si existe un subespacio unidimensional* W *de* V *tal que* $T(x) = -x$ *para toda* $x \in W$ *y* $T(y) = y$ *para toda* $y \in W^\perp$. *Dentro de este contexto* T *se llama* reflexión de V alrededor de $W^\perp$.

Debería hacerse notar que las rotaciones y las reflexiones (o aun las composiciones) son operadores ortogonales. (Ver el Ejercicio 2.) La finalidad principal de esta sección es establecer que la recíproca también es cierta, esto es, que cualquier operador ortogonal en un espacio real con producto interior dimensionalmente finito es el resultado de la composición de rotaciones y reflexiones.

Ejemplo 29. Caracterización de operadores ortogonales en un espacio real unidimensional con producto interior.

Sea T un operador ortogonal en un espacio unidimensional con producto interior V. Tómese cualquier vector x no nulo en V. Entonces $V = L(\{x\})$ y así $T(x) = \lambda x$ para alguna $\lambda \in R$. Como T es ortogonal y λ es un eigenvalor de T, $\lambda = \pm 1$. Si $\lambda = 1$, entonces T es la identidad en V y, por lo tanto, T es una rotación. Si $\lambda = -1$, entonces $T(x) = -x$ para toda $x \in V$ y, por lo tanto, T es una reflexión de V alrededor de $V^\perp = \{0\}$. Luego, T es una rotación o una reflexión. Nótese que en el primer caso $\det(T) = 1$ y en el segundo caso $\det(T) = -1$.

Ejemplo 30. Algunas reflexiones típicas.

(a) Sea T: $R^2 \to R^2$ definida mediante $T(a, b) = (-a, b)$. Si $W = L(\{e_1\})$, entonces $T(x) = -x$ para toda $x \in W$ y $T(y) = y$ para toda $y \in W^\perp$. Luego, T es una reflexión de R^2 alrededor de $W^\perp = L(\{e_2\})$, el eje y.

(b) Sea T: $R^3 \to R^3$ definida mediante $T(a, b, c) = (a, b, -c)$. Si $W = L(\{e_3\})$, entonces $T(x) = -x$ para toda $x \in W$ y $T(y) = y$ para toda $y \in W^\perp = L(\{e_1, e_2\})$, el plano xy.

En el Ejemplo 29 caracterizamos a todos los operadores ortogonales en un espacio real unidimensional con producto interior. El teorema siguiente caracteriza a todos los operadores ortogonales en un espacio real bidimensional con producto interior. La demostración de este resultado se obtiene fácilmente del Ejercicio 20 de la Sección 7.7, pues una reflexión alrededor del eje x seguida de una rotación por θ es una reflexión alrededor de la recta que pasa por el origen con una pendiente de $\tan \frac{1}{2}\theta$.

Teorema 7.22. *Sea* T *un operador ortogonal en un espacio real bidimensional con producto interior* V. *Entonces,* T *es o una rotación o una reflexión. Además,* T *es una rotación si y sólo si* $\det(T) = 1$ *y* T *es una reflexión si y sólo si* $\det(T) = -1$.

De acuerdo con la definición, cualquier reflexión en R^2 tiene los eigenvalores 1 y -1, y cualquier par de eigenvectores correspondientes a estos eigenvalores son ortogonales. Además, el eigenespacio de T correspondiente a $\lambda = 1$ es unidimensional y, por lo tanto, puede ser descrito

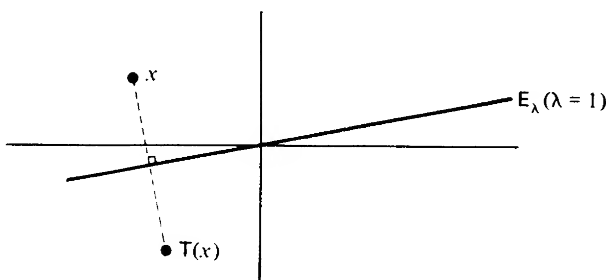


figura 7.5

como una recta que pasa por el origen. Geométricamente, T refleja a todos los puntos de R^2 alrededor de esta recta. (Véase Fig. 7.5.) Por ejemplo, si

$$A = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{-1}{\sqrt{5}} \end{pmatrix},$$

es claro que L_A es un operador ortogonal en R^2 y que $\det(L_A) = \det(A) = -1$. Por lo tanto, L_A es una reflexión de acuerdo con el Teorema 7.22. Para encontrar el subespacio sobre el cual se refleja L_A es suficiente encontrar un eigenvector de L_A correspondiente al eigenvalor $\lambda = 1$. Uno de tales eigenvectores es

$$x = \begin{pmatrix} \frac{1}{\sqrt{5}} + 1 \\ \frac{2}{\sqrt{5}} \end{pmatrix}.$$

En consecuencia, el subespacio sobre el cual L_A se refleja es la recta

$$\left\{ t\begin{pmatrix} \frac{1}{\sqrt{5}} + 1 \\ \frac{2}{\sqrt{5}} \end{pmatrix} : \quad t \in R \right\}.$$

Corolario. *Sea* V *un espacio real bidimensional con producto interior. La composición de una reflexión y una rotación en* V *es una reflexión en* V.

DEMOSTRACIÓN. Si T_1 es una reflexión en V y T_2 es una rotación en V, entonces, de acuerdo con el Teorema 7.22, $\det(T_1) = 1$ y $\det(T_2) = -1$. Sea $T = T_2T_1$ la composición. Como T_2 y T_1 son ortogonales, también T lo es. Además, $\det(T) = \det(T_2) \cdot \det(T_1) = -1$. Luego, por el Teorema 7.22, T es una reflexión. La demostración para T_1T_2 es semejante. ■

Estudiaremos ahora operadores ortogonales en espacios de dimensión superior.

Lema. *Si* T *es un operador lineal en un espacio real con producto interior dimensionalmente finito y no nulo* V, *entonces existe un subespacio* T-*invariante* W *de* V *tal que* $1 \leq \dim(W) \leq 2$.

DEMOSTRACIÓN. Fíjese una base ordenada $\beta = \{x_1, x_2, \ldots, x_n\}$ para V, y sea $A = [T]_\beta$. Sea ϕ_β: $V \to R^n$ la transformación lineal definida mediante $\phi_\beta(x_i) = e_i$ para $i = 1, 2, \ldots, n$. Entonces ϕ_β es un isomorfismo, y hemos visto en la Sección 2.4 que el diagrama de la figura 7.6 es conmutativo, esto es que $L_A\phi_\beta = \phi_\beta T$. En consecuencia, es suficiente demostrar que existe un subespacio Z de R^n L_A-invariante tal que $1 \leq \dim(Z) \leq 2$. Si entonces definimos a $W = \phi_\beta^{-1}(Z)$, se tendrá que W satisface la conclusión del teorema. (Véase Ejercicio 12.)

Puede considerarse a la matriz A como una matriz de $n \times n$ sobre C y como tal puede ser utilizada para definir un operador lineal U en C^n mediante $U(x) = Ax$ para todos los vectores columna x en C^n. Como U es un operador en un espacio vectorial dimensionalmente finito sobre C,

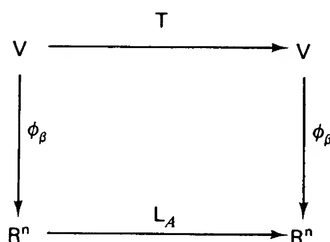


figura 7.6

tiene un eigenvalor $\lambda \in C$. Sea $x \in C^n$ un eigenvector correspondiente a λ. Podemos escribir $\lambda = \lambda_1 + i\lambda_2$, donde λ_1 y λ_2 son reales y

$$x = \begin{pmatrix} a_1 + ib_1 \\ a_2 + ib_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n + ib_n \end{pmatrix},$$

donde las a_i y las b_i son reales. Luego, haciendo a

$$x_1 = \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{pmatrix} \quad \text{y} \quad x_2 = \begin{pmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{pmatrix},$$

tenemos que $x = x_1 + ix_2$, donde x_1 y x_2 son n-dimensionales formadas por elementos reales. Nótese que al menos una de x_1 o x_2 es no nula ya que $x \neq 0$. Por lo tanto,

$$\begin{aligned} U(x) &= \lambda x = (\lambda_1 + i\lambda_2)(x_1 + ix_2) \\ &= (\lambda_1 x_1 - \lambda_2 x_2) + i(\lambda_1 x_2 + \lambda_2 x_1). \end{aligned} \tag{14}$$

De la misma manera

$$U(x) = A(x_1 + ix_2) = Ax_1 + iAx_2. \tag{15}$$

Comparando las partes real e imaginaria de las Ecuaciones (14) y (15), concluimos que

$$Ax_1 = \lambda_1 x_1 - \lambda_2 x_2 \quad \text{y} \quad Ax_2 = \lambda_1 x_1 + \lambda_2 x_1. \tag{16}$$

Finalmente, sea $Z = L(\{x_1, x_2\})$ tomándolo como subespacio de R^n. Como $x_1 \neq 0$ o $x_2 \neq 0$, Z es no nulo. Luego, $1 \leq \dim(Z) \leq 2$ y por la Ecuación (16) Z es L_A-invariante. ■

Teorema 7.23. *Sea* T *un operador ortogonal en un espacio real no nulo con producto interior dimensionalmente finito* V. *Entonces, existe una colección de subespacios* T*-invariantes ortogonales por parejas* $\{W_1, W_2, \ldots, W_m\}$ *de* V *tales que*

(a) $1 \leq dim(W_i) \leq 2$ *para* $i = 1, 2, \ldots, m$.
(b) $V = W_1 \oplus W_2 \oplus \cdots \oplus W_m$.

DEMOSTRACIÓN. La demostración se hará por inducción sobre dim(V). Si dim(V) = 1, el resultado es evidente. Por tanto, supóngase que el resultado es cierto para $\dim(V) < n$ para algún entero fijo $n > 1$.

Supóngase que $\dim(V) = n$. En virtud del lema existe un subespacio W_1 de V T-invariante tal que $1 \leq \dim(W_1) \leq 2$. Si $W_1 = V$, el resultado queda establecido. De lo contrario, $W_1^\perp \neq \{0\}$. En virtud del Ejercicio 13 $W_1^\perp$ es T-invariante, y la restricción de T a $W_1^\perp$ es ortogonal. Como $\dim(W_1^\perp) < n$, podemos aplicar la hipótesis de inducción a $T_{W_1^\perp}$ y concluir que existe una colección de subespacios T-invariantes ortogonales por parejas $\{W_2, W_3, \ldots, W_m\}$ de $W_1^\perp$ tales que $1 \leq \dim(W_i) \leq 2$ para $i = 2, 3, \ldots, m$ y $W_1^\perp = W_2 \oplus W_3 \oplus \cdots \oplus W_m$. Por lo tanto, $\{W_1, W_2, \ldots, W_m\}$ es ortogonal por parejas y

$$V = W_1 \oplus W_1^\perp = W_1 \oplus W_2 \oplus \cdots \oplus W_m. \quad \blacksquare$$

Aplicando el Ejemplo 29 y el Teorema 7.22 en el contexto del Teorema 7.23, podemos concluir que la restricción de T a W_i es una rotación o bien una reflexión para cada $i = 1, 2, \ldots, m$. Así pues, en cierto sentido, T está formado de rotaciones y reflexiones. Desafortunadamente, puede decirse muy poco sobre la descomposición de V en el Teorema 7.23 en términos de unicidad. Por ejemplo, las W_i, el número m de W_i y el número de W_i para las que T_{W_i} es una reflexión no son únicas. Aun cuando no es único el número de W_i para los cuales T_{W_i} es una reflexión, el que este número sea par o impar es una propiedad intrínseca de T. Además, siempre podemos descomponer a V de manera que T_{W_i} sea una reflexión para a lo más un W_i. Estos hechos se establecen en el resultado siguiente.

Teorema 7.24. *Sean* $T, V, W_1, \ldots, W_m$ *como en el Teorema 7.23.*

(a) *El que el número de* i *para las cuales* T_{W_i} *es una reflexión sea par o impar depende de que* $det(T) = 1$ *o* $det(T) = -1$.
(b) *Siempre es posible descomponer a* V *como en el Teorema 7.23 de manera que el número de* i *para las que* T_{W_i} *es una reflexión sea cero o uno, dependiendo de que* $det(T) = 1$ *o* $det(T) = -1$. *Además, si* T_{W_i} *es una reflexión, entonces* $dim(W_i) = 1$.

DEMOSTRACIÓN.

(a) Sea r el número de W_i en la descomposición para las cuales T_{W_i} es una reflexión. Por lo tanto, de acuerdo con el Ejercicio 14,

$$\det(T) = \det(T_{W_1})\cdot\det(T_{W_2})\cdot\cdots\cdot\det(T_{W_m}) = (-1)^r,$$

con lo que se demuestra el inciso (a).

(b) Sea $E = \{x \in V\colon\ T(x) = -x\}$; entonces E es un subespacio T-invariante de V. Si $W = E^{\perp}$, entonces W es T-invariante. Así, aplicando el Teorema 7.23 a T_W obtenemos una colección de subespacios T-invariantes ortogonales por parejas $\{W_1, W_2, \ldots, W_k\}$ de W tales que $1 \leq \dim(W_i) \leq 2$ para $1 \leq i \leq k$ y $W = W_1 \oplus W_2 \oplus \cdots \oplus W_k$. Obsérvese que, para cada $i = 1, 2, \ldots, k$, T_{W_i} es una rotación. De lo contrario, si T_{W_i} es una reflexión, existe un elemento $x \in W_i$ no nulo, para el cual $T(x) = -x$. Pero entonces $x \in W_i \cap E \subseteq E^{\perp} \cap E = \{0\}$, lo cual es una contradicción. Si $E = \{0\}$, se obtiene el resultado. De lo contrario, tómese una base ortonormal β para E que contenga p elementos $(p > 0)$. Es posible descomponer a β en una unión disjunta por parejas $\beta = \beta_1 \cup \cup \beta_2 \cup \ldots \cup \beta_r$ tal que cada β_i contenga exactamente dos elementos para $i < r$ y que β_r contega dos elementos cuando p sea par y un elemento si p es impar. Para cada $i = 1, 2, \ldots, r$, sea $W_{k+i} = L(\beta_i)$. Entonces, claramente $\{W_1, W_2, \ldots, W_k, \ldots, W_{k+r}\}$ es ortogonal por parejas y

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_k \oplus \cdots \oplus W_{k+r}. \tag{17}$$

Además, si cualquier β_i contiene dos elementos, entonces

$$\det(T_{W_{k+i}}) = \det([T_{W_{k+i}}]_{\beta_i}) = \det\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = 1.$$

Entonces $T_{W_{k+i}}$ es una rotación y, por lo tanto, T_{W_j} es una rotación para $j < k + r$. Si β_r está formado por un elemento, entonces $\dim(W_{k+r}) = 1$ y $\det(T_{W_{k+r}}) = \det([T_{W_{k+r}}]_{\beta_r}) = \det(-1) = -1$. Así pues, $T_{W_{k+r}}$ es una reflexión, por el Teorema 7.23, y concluimos que la descomposición de la Ecuación (17) satisface la condición del inciso (b). ■

Como consecuencia del teorema anterior, un operador ortogonal se puede descomponer como un producto de rotaciones y reflexiones.

Corolario. *Sea* T *un operador ortogonal en un espacio real con producto interior dimensionalmente finito* V. *Entonces existe una colección* $\{T_1, T_2, \ldots, T_m\}$ *de operadores ortogonales en* V *tales que*

(a) *Para cada* i, T_i *es una reflexión o bien una rotación.*
(b) *Para una* i *como máximo,* T_i *es una reflexión.*
(c) $T_iT_j = T_jT_i$ *para toda* i *y toda* j.
(d) $T = T_1T_2 \ldots T_m$.
(e) $$\det(T) = \begin{cases} 1 & \text{si } T_i \text{ es una rotación para cada } i \\ -1 & \text{en cualquier otro caso} \end{cases}$$

DEMOSTRACIÓN. Como en la demostración del inciso (b) del Teorema 7.24 podemos escribir $V = W_1 \oplus W_2 \oplus \cdots \oplus W_m$, donde T_{W_i} es una rotación para $i < m$. Para cada $i = 1, 2, \ldots, m$, defínase a T_i: $V \to V$ mediante

$$T_i(x_1 + \cdots + x_m) = x_1 + \cdots + x_{i-1} + T(x_i) + x_{i+1} + \cdots + x_m,$$

donde $x_j \in W_j$ para toda j. Puede demostrarse fácilmente que cada T_i es un operador ortogonal en V. De hecho, T_i es una rotación o una reflexión dependiendo de que T_{W_i} sea una rotación o una reflexión. Esto demuestra los incisos (a) y (b). Las demostraciones de (c), (d) y (e) se dejan como ejercicio. (Véase Ejercicio 15.)

Ejemplo 31. Operadores ortogonales en un espacio real tridimensional con producto interior.

Sea T un operador ortogonal en un espacio real tridimensional con producto interior V. Demostraremos que T puede descomponerse en una rotación y a lo más una reflexión. Sea $V = W_1 \oplus W_2 \oplus \cdots \oplus W_m$ una descomposición como la del Teorema 7.24(b). Claramente $m = 2$ o $m = 3$.

Si $m = 2$, entonces $V = W_1 \oplus W_2$. Sin pérdida de generalidad, supóngase que $\dim(W_1) = 1$ y $\dim(W_2) = 2$. Luego, T_{W_1} es una reflexión, o bien la identidad en W_1, y T_{W_2} es una rotación. Definiendo T_1 y T_2 como en la demostración del corolario del Teorema 7.24, tenemos que $T = T_1T_2$ es la composición de una rotación y a lo más una reflexión. (Nótese que si T_{W_1} no es una reflexión, entonces T_1 es la identidad en V y $T = T_2$.)

Si $m = 3$, entonces $V = W_1 \oplus W_2 \oplus W_3$ y $\dim(W_i) = 1$ para toda i. Para cada i, sea T_i como en la demostración del corolario del Teorema 7.24. Si T_{W_i} no es una reflexión, entonces T_i es la identidad en W_i. De lo contrario T_i es una reflexión. Como T_{W_i} es una reflexión para una i como máximo, concluimos que T es una reflexión sencilla o bien la identidad (una rotación).

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas. Supóngase para lo que sigue que los espacios vectoriales subyacentes son espacios reales con producto interior dimensionalmente finitos.
   (a) Cualquier operador ortogonal es una rotación o una reflexión.
   (b) La composición de dos rotaciones cualesquiera en un espacio bidimensional es una rotación.
   (c) La composición de dos rotaciones cualesquiera en un espacio de tres dimensiones es una rotación.

(d) La composición de dos rotaciones cualesquiera en un espacio de cuatro dimensiones es una rotación.
(e) El operador identidad es una rotación.
(f) La composición de dos reflexiones es una reflexión.
(g) Cualquier operador ortogonal es una composición de rotaciones.
(h) Para cualquier operador ortogonal T, si $\det(T) = -1$, entonces T es una reflexión.
(i) Las reflexiones siempre tienen eigenvalores.
(j) Las rotaciones siempre tienen eigenvalores.

2. Demostrar que las rotaciones, las reflexiones y las composiciones de rotaciones y reflexiones son operadores ortogonales.

3. Sea

$$A = \begin{pmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} \quad \text{y} \quad B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

(a) Demostrar que L_A: $R^2 \to R^2$ es una reflexión.
(b) Encontrar el eje en R^2 alrededor del cual se refleja L_A, esto es, el subespacio de R^2 en el cual L_A actúa como la identidad.
(c) Demostrar que L_{AB} y L_{BA} son rotaciones.

4. Para cualquier número real ϕ, sea

$$A = \begin{pmatrix} \cos\phi & \operatorname{sen}\phi \\ \operatorname{sen}\phi & -\cos\phi \end{pmatrix}.$$

(a) Demostrar que L_A es una reflexión.
(b) Encontrar el eje en R^2 alrededor del cual L_A se refleja.

5. Para cualquier número real ϕ, definir a $T_\phi = L_A$, donde

$$A = \begin{pmatrix} \cos\phi & -\operatorname{sen}\phi \\ \operatorname{sen}\phi & \cos\phi \end{pmatrix}.$$

(a) Demostrar que cualquier rotación de R^2 es de la forma T_ϕ para alguna ϕ.
(b) Demostrar que $T_\phi T_\psi = T_{(\phi+\psi)}$ para cualquier ϕ, $\psi \in R$.
(c) Deducir que cualquier par de rotaciones en R^2 conmutan.

6. Demostrar que la composición de cualquier par de rotaciones en R^3 es una rotación en R^3.

7. Dados los números reales ϕ y ψ, defínanse matrices

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\operatorname{sen}\phi \\ 0 & \operatorname{sen}\phi & \cos\phi \end{pmatrix} \quad \text{y} \quad B = \begin{pmatrix} \cos\psi & -\operatorname{sen}\psi & 0 \\ \operatorname{sen}\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

(a) Demostrar que L_A y L_B son rotaciones.
(b) Demostrar que L_{AB} es una rotación.
(c) Encontrar el eje de rotación para L_{AB}.

8. Demostrar que ningún operador ortogonal puede ser a la vez una rotación y una reflexión.

9. Demostrar que si V es un espacio real de dos o tres dimensiones con producto interior, entonces la composición de dos reflexiones en V es una rotación en V.

10. Dar un ejemplo de un operador ortogonal que no sea ni reflexión ni rotación.

11. Sea V un espacio real con producto interior dimensionalmente finito. Defínase a $T: V \rightarrow V$ mediante $T(x) = -x$. Demostrar que T es un producto de rotaciones si y sólo si dim(V) es par.

12. Completar la demostración del lema del Teorema 7.23, demostrando que $W = \phi_\beta^{-1}(Z)$ satisface las condiciones requeridas.

13. Sea T un operador ortogonal [unitario] en un espacio real [complejo] dimensionalmente finito con producto interior V. Si W es un subespacio T-invariante de V, demostrar que

(a) T_W es un operador ortogonal [unitario] en W.
(b) $W^\perp$ es un subespacio T-invariante de V.
Sugerencia: Utilizar el hecho de que T_W es uno-a-uno y sobreyectivo para llegar a la conclusión de que, para cualquier $y \in W$, $T^*(y) = T^{-1}(y) \in W$.
(c) $T_{W^\perp}$ es un operador ortogonal [unitario] en W.

14. Sea T un operador lineal en un espacio vectorial dimensionalmente finito V. Supóngase que V es una suma directa de subespacios T-invariantes $V = W_1 \oplus W_2 \oplus \cdots \oplus W_k$. Demostrar que $\det(T) = \det(T_{W_1}) \cdot \det(T_{W_2}) \cdot \cdots \cdot \det(T_{W_k})$.

15. Completar la demostración del corolario del Teorema 7.24.

16. Sea T un operador ortogonal en un espacio real n-dimensional con producto interior V. Supóngase que T no es la identidad. Demostrar que

(a) Si n es impar, T puede expresarse como la composición de a lo más una reflexión y a lo más $\frac{1}{2}(n - 1)$ rotaciones.
(b) Si n es par, entonces T puede expresarse como la composición de a lo más $\frac{1}{2}n$ rotaciones o como la composición de una reflexión y a lo más $\frac{1}{2}(n - 2)$ rotaciones.

17. Sea V un espacio real con producto interior de dimensión 2. Para cualesquier x, $y \in V$ tales que $x \neq y$ y $||x|| = ||y|| = 1$, demostrar que existe una rotación única T en V, tal que $T(x) = y$.

7.9 PROYECCIONES ORTOGONALES Y EL TEOREMA ESPECTRAL

En esta sección nos basaremos en gran parte en el Teorema 7.17 para desarrollar una elegante representación de un operador normal T en un espacio complejo dimensionalmente finito con producto interior. Demostraremos que tal operador puede escribirse en la forma $\lambda_1 T_1 + \ldots + \lambda_k T_k$, donde $\lambda_1, \ldots, \lambda_k$ son los distintos eigenvalores de T y $T_1, \ldots, T_k$ son "proyecciones ortogonales". Pero primero debemos desarrollar algunos resultados sobre estas proyecciones especiales.

El lector recordará de la Sección 2.1 que una transformación lineal $T: V \to V$ es una proyección (en su rango $R(T)$) si $V = R(T) \oplus N(T)$. De hecho, T es una proyección si y sólo si $T = T^2$ (véase Ejercicio 14 de la Sección 2.3).

Definición. *Sea* V *un espacio con producto interior, y sea* $T: V \to V$ *una proyección. Decimos que* T *es una* proyección ortogonal *si* $R(T)^\perp = N(T)$ *y* $N(T)^\perp = R(T)$.

Nótese que, por el Ejercicio 12(c) de la Sección 7.2, si V es dimensionalmente finito, sólo tenemos que suponer que una de las condiciones anteriores se cumple. Por ejemplo, si $R(T)^\perp = N(T)$, entonces $R(T) = R(T)^{\perp\perp} = N(T)^\perp$.

Ahora supongamos que W es un subespacio dimensionalmente finito de un espacio con producto interior V. El Teorema 7.6 garantiza que existe una proyección ortogonal en W. Podemos decir aún más —existe exactamente una proyección ortogonal en W. Ya que si T y U son proyecciones ortogonales en W, entonces $R(T) = W = R(U)$; por lo tanto $N(T) = R(T)^\perp = R(U)^\perp = N(U)$, y como todas las proyecciones están determinadas de manera única por su rango y por su espacio nulo (kernel), tenemos que $T = U$. Llamamos a T la *proyección ortogonal sobre* W. Para comprender la diferencia geométrica entre una proyección arbitraria sobre W y la proyección ortogonal sobre W, sean $V = R^2$ y $W = L\{(1, 1)\}$. Defínanse U y T como en la Figura 7.7, donde $T(v)$ es el pie de una perpendicular que parte de v e intersecta a la recta $y = x$, y $U(a_1, a_2) = (a_1, a_1)$. Entonces T es la proyección ortogonal sobre W, y U es una proyección sobre W que no es ortogonal. Nótese que $v - T(v) \in W^\perp$, mientras que $v - U(v) \notin W^\perp$.

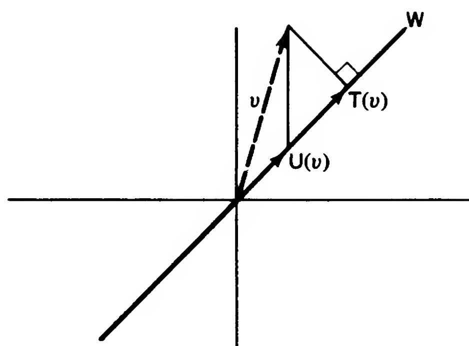


figura 7.7

De la Figura 7.7 vemos que $T(v)$ es la "mejor aproximación en W para v"; esto es, si $w \in W$, $||w - v|| \geq ||T(v) - v||$. Esta propiedad de aproximación caracteriza a T. De hecho, muchos autores definen las proyecciones ortogonales en términos de esta propiedad.

Teorema 7.25. *Sea* W *un subespacio dimensionalmente finito de un espacio con producto interior* V, *y sea* T *la proyección ortogonal sobre* W. *Entonces, para cualquier* $v \in W$: *el vector* $T(v)$ *es el único elemento de* W *que se acerca más a* v; *esto es,* $||v - T(v)|| \leq ||v - w||$ *para toda* $w \in W$.

DEMOSTRACIÓN. Sea $v \in V$. Como T es una proyección ortogonal, podemos escribir $v = T(v) + (v - T(v))$, donde $v - T(v) \in N(T) = W^{\perp}$. Sea $w \in W$. En virtud del Ejercicio 10 de la Sección 7.1, tenemos que

$$\begin{aligned} ||v - w||^2 &= ||T(v) - w + (v - T(v))||^2 \\ &= ||T(v) - w||^2 + ||v - T(v)||^2 \geq ||v - T(v)||^2, \end{aligned}$$

demostrando así la desigualdad anterior. Si para alguna $w \in W$ tenemos que $||v - w|| = ||v - T(v)||$, vemos del cálculo anterior que $||T(v) - w||^2 = 0$; esto es, $w = T(v)$. ■

En la Sección 7.10 veremos una aplicación muy importante del Teorema 7.25 al tema de la aproximación por mínimos cuadrados que aparece frecuentemente en estadística.

Por ahora, aplicaremos el Teorema 7.25 para obtener un resultado muy conocido en el Análisis de Fourier. Recuérdese el espacio con producto interior H de las funciones continuas en el intervalo $[0, 2\pi]$ introducido en la Sección 7.1. Defínase a un *polinomio trigonométrico de grado* n como una función $g \in H$ de la forma

$$g(x) = \sum_{j=-n}^{n} a_j e^{ijx},$$

donde a_n o a_{-n} son diferentes de cero.

Sea $f \in \mathsf{H}$. Demostraremos que la mejor aproximación de f por un polinomio trigonométrico de grado menor o igual a n es el polinomio cuyos coeficientes son los coeficientes de Fourier de f con respecto al conjunto ortonormal $\{e^{ijx}: j \text{ es un entero}\}$.

Para este resultado, sea $\mathsf{W} = L(\{e^{ijx}: |j| \leq n\})$ y sea T la proyección ortogonal sobre W. El Teorema 7.25 nos dice que

$$\mathsf{T}(f) = \sum_{j=-n}^{n} (f, e^{ijx})e^{ijx}$$

es la mejor representación de f en H. (Véase también el Corolario 1 del Teorema 7.6.)

Una caracterización algebraica de proyecciones ortogonales se tiene en el teorema siguiente.

Teorema 7.26. *Sea* V *un espacio con producto interior, y sea* T *un operador lineal en* V. *Entonces* T *es una proyección ortogonal si y sólo si* $\mathsf{T}^2 = \mathsf{T} = \mathsf{T}^*$.

DEMOSTRACIÓN. Supóngase que T es una proyección ortogonal. Como $\mathsf{T} = \mathsf{T}^2$ por el hecho de ser T una proyección, sólo necesitamos demostrar que $\mathsf{T} = \mathsf{T}^*$. Ahora bien, $\mathsf{V} = \mathsf{R}(\mathsf{T}) \oplus \mathsf{N}(\mathsf{T})$ y $\mathsf{R}(\mathsf{T})^{\perp} = \mathsf{N}(\mathsf{T})$. Si $x, y \in \mathsf{V}$, entonces $x = x_1 + x_2$ e $y = y_1 + y_2$, donde $x_1, y_1 \in \mathsf{R}(\mathsf{T})$ y $x_2, y_2 \in \mathsf{N}(\mathsf{T})$ Por lo tanto

$$(x, \mathsf{T}(y)) = (x_1 + x_2, y_1) = (x_1, y_1) + (x_2, y_1) = (x_1, y_1)$$

y

$$(x, \mathsf{T}^*(y)) = (\mathsf{T}(x), y) = (x_1, y_1 + y_2) = (x_1, y_1) + (x_1, y_2) = (x_1, y_1).$$

Así, $(x, \mathsf{T}(y)) = (x, \mathsf{T}^*(y))$ para todo $x, y \in \mathsf{V}$ y, por lo tanto, $\mathsf{T} = \mathsf{T}^*$.

Ahora supóngase que $\mathsf{T} = \mathsf{T}^2 = \mathsf{T}^*$. Entonces, por el Ejercicio 14 de la Sección 2.3, T es una proyección y, por lo tanto, debemos demostrar que $\mathsf{R}(\mathsf{T}) = \mathsf{N}(\mathsf{T})^{\perp}$ y $\mathsf{R}(\mathsf{T})^{\perp} = \mathsf{N}(\mathsf{T})$. Sean $x \in \mathsf{R}(\mathsf{T})$ y $y \in \mathsf{N}(\mathsf{T})$. Entonces $x = \mathsf{T}(x) = \mathsf{T}^*(x)$, y así $(x, y) = (\mathsf{T}^*(x), y) = (x, \mathsf{T}(y)) = (x, 0) = 0$. Por lo tanto $x \in \mathsf{N}(\mathsf{T})^{\perp}$, de donde se tiene que $\mathsf{R}(\mathsf{T}) \subseteq \mathsf{N}(\mathsf{T})^{\perp}$.

Sea $y \in \mathsf{N}(\mathsf{T})^{\perp}$. Debemos demostrar que $y \in \mathsf{R}(\mathsf{T})$, esto es, que $\mathsf{T}(y) = y$. Ahora bien,

$$\begin{aligned} \|y - \mathsf{T}(y)\|^2 &= (y - \mathsf{T}(y), y - \mathsf{T}(y)) \\ &= (y, y - \mathsf{T}(y)) - (\mathsf{T}(y), y - \mathsf{T}(y)). \end{aligned}$$

Como $y - \mathsf{T}(y) \in \mathsf{N}(\mathsf{T})$, el primer término es cero. Pero también $(\mathsf{T}(y), y - \mathsf{T}(y)) = (y, \mathsf{T}^*(y - \mathsf{T}(y))) = (y, \mathsf{T}(y - \mathsf{T}(y))) = (y, 0) = 0$. Así, tenemos que $y - \mathsf{T}(y) = 0$; esto es, $y = \mathsf{T}(y) \in \mathsf{R}(\mathsf{T})$. Por lo tanto $\mathsf{R}(\mathsf{T}) = \mathsf{N}(\mathsf{T})^{\perp}$.

Utilizando lo anterior, tenemos que $\mathsf{R}(\mathsf{T})^{\perp} = \mathsf{N}(\mathsf{T})^{\perp\perp} \supseteq \mathsf{N}(\mathsf{T})$ (en virtud del Ejercicio 12(b) de la Sección 7.2). Unicamente necesitamos demostrar que si $x \in \mathsf{R}(\mathsf{T})^{\perp}$, entonces $x \in \mathsf{N}(\mathsf{T})$. Para cualquier $y \in \mathsf{V}$, tene-

mos que $(T(x), y) = (x, T^*(y)) = (x, T(y)) = 0$. Así, $T(x) = 0$ y por tanto $x \in N(T)$. ■

Sea V un espacio dimensionalmente finito con producto interior, W un subespacio de V y T la proyección ortogonal en W. Podemos escoger una base ortonormal $\beta = \{x_1, \ldots, x_n\}$ para V tal que $\{x_1, \ldots, x_k\}$ sea una base para W. Entonces $[T]_\beta$ es una matriz diagonal con unos a lo largo de los primeros k elementos de la diagonal y ceros en cualquier otra posición. De hecho, $[T]_\beta$ tiene la forma

$$\begin{pmatrix} I_k & O_1 \\ O_2 & O_3 \end{pmatrix},$$

donde O_1, O_2 y O_3 son matrices nulas.

Si U es una proyección cualquiera sobre W, podemos escoger una base γ para V tal que $[U]_\gamma$ tenga la forma anterior; sin embargo, no es necesario que γ sea ortonormal.

Teorema 7.27. *(El teorema espectral.) Supóngase que* T *es un operador lineal en un espacio con producto interior dimensionalmente finito* V *sobre* F. *Supóngase que* T *es normal si* $F = C$ *y que* T *es autoadjunta si* $F = R$. *Si* $\lambda_1, \ldots, \lambda_k$ *son los distintos eigenvalores de* T, *sea* $W_i = E_{\lambda_i} = \{x \in V: T(x) = \lambda_i x\}$ *el eigenespacio de* T *correspondiente al eigenvalor* $\lambda_i (1 \leq i \leq k)$ *y sea* T_i *la proyección ortogonal sobre* $W_i (1 \leq i \leq k)$. *Entonces*

(a) $V = W_1 \oplus \ldots \oplus W_k$.
(b) *Si* W'_i *es la suma directa de los subespacios* W_j, $j \neq i$, *entonces* $W_i^\perp = W'_i$.
(c) $T_i T_j = \delta_{ij} T_i$ *para* $1 \leq i, j \leq k$.
(d) $I = T_1 + \ldots + T_k$.
(e) $T = \lambda_1 T_1 + \ldots + \lambda_k T_k$.

DEMOSTRACIÓN.

(a) De acuerdo con el Teorema 7.17, T es diagonalizable y entonces, de acuerdo con el Teorema 5.14, $V = W_1 \oplus \ldots \oplus W_k$.

(b) Si $x \in W_i$ y $y \in W_j$ para algunas i y j, entonces $(x, y) = 0$, de acuerdo con el Teorema 7.15. Se infiere de esto fácilmente que $W'_i \subseteq W_i^\perp$. Ahora bien, de (a) tenemos que

$$\dim(W'_i) = \sum_{j \neq i} \dim(W_j) = \dim(V) - \dim(W_i).$$

Por otra parte, de acuerdo con el Corolario 2 del Teorema 7.6, tenemos que $\dim(W_i^\perp) = \dim(V) - \dim(W_i)$. Por lo tanto, $W'_i = W_i^\perp$, con lo que se demuestra el inciso (b).

La demostración del inciso (c) se deja como ejercicio.

Como T_i es la proyección ortogonal sobre W_i, tenemos de (b) que $N(T_i) = R(T_i)^{\perp} = W_i^{\perp} = W_i'$. Por lo tanto, para $x \in V$ tenemos que $x = x_1 + \ldots + x_k$, donde $x_j \in W_j$ y $T_i(x) = x_i$, demostrando el inciso (d).

(e) Para $x \in V$, escríbase $x = x_1 + \ldots + x_k$, donde $x_j \in W_j (1 \leq j \leq k)$. Entonces $T(x) = T(x_1) + \ldots + T(x_k) = \lambda_1 x_1 + \ldots + \lambda_k x_k = \lambda_1 T_1(x) + \ldots + \lambda_k T_k(x) = (\lambda_1 T_1 + \ldots + \lambda_k T_k)(x)$. ■

El conjunto $\{\lambda_1, \ldots, \lambda_k\}$ de eigenvalores de T se llama *espectro de* T, la suma $I = T_1 + \ldots + T_k$ del inciso (d) se llama *resolución del operador identidad inducida por* T, y la suma $T = \lambda_1 T_1 + \ldots + \lambda_k T_k$ del inciso (e) se denomina *descomposición espectral de* T. Como los distintos eigenvalores de T quedan determinados de manera única (hasta el orden) por los subespacios W_i (y, por lo tanto, por las proyecciones ortogonales T_i), la descomposición espectral de T es única.

Con la notación anterior, sea β la unión de las bases ortonormales de los W_i y sea $m_i = \dim(W_i)$. (Luego, m_i es la multiplicidad de λ_i.) Entonces, $[T]_\beta$ tiene la forma

$$\begin{pmatrix} \lambda_1 I_{m_1} & O & \cdots & O \\ O & \lambda_2 I_{m_2} & \cdots & O \\ \vdots & \vdots & & \vdots \\ O & O & \cdots & \lambda_k I_{m_k} \end{pmatrix};$$

esto es, $[T]_\beta$ es una matriz diagonal en donde los elementos de la diagonal son los eigenvalores λ_i de T, y cada λ_i se repite m_i veces. Si $T = \lambda_1 T_1 + \ldots + \lambda_k T_k$ como en el inciso (e) del teorema espectral, entonces se tiene (del Ejercicio 7) que $g(T) = g(\lambda_1)T_1 + \ldots + g(\lambda_k)T_k$ para cualquier polinomio g. Más adelante utilizaremos este hecho.

Enunciaremos ahora algunos corolarios interesantes del teorema espectral; muchos resultados más se encuentran en los ejercicios. Para lo que sigue supondremos que V es un espacio con producto interior dimensionalmente finito sobre F y que T es un operador lineal en V.

Corolario 1. *Si* $F = C$, *entonces* T *es normal si y sólo si* $T^* = g(T)$ *para algún polinomio* g.

DEMOSTRACIÓN. Supóngase primero que T es normal. Sea $T = \lambda_1 T_1 + \ldots + \lambda_k T_k$ la descomposición espectral de T. Tomando el adjunto de ambos lados de la ecuación anterior tenemos $T^* = \bar{\lambda}_1 T_1 + \ldots + \bar{\lambda}_k T_k$ ya que cada uno de los T_i son autoadjuntos. Empleando la fórmula de interpolación de Lagrange (ver p. 49), podemos escoger un polinomio g tal que $g(\lambda_i) = \bar{\lambda}_i$ para $1 \leq i \leq k$. Entonces

$$\begin{aligned} g(T) &= g(\lambda_1)T_1 + \ldots + g(\lambda_k)T_k \\ &= \bar{\lambda}_1 T_1 + \ldots + \bar{\lambda}_k T_k \\ &= T^*. \end{aligned}$$

Recíprocamente, si $T^* = g(T)$ para algún polinomio g, entonces T conmuta con T^*, puesto que T conmuta con todo polinomio en T. ■

Corolario 2. *Si $F = C$, entonces T es unitario si y sólo si T es normal y $|\lambda| = 1$ para todo eigenvalor λ de T.*

DEMOSTRACIÓN. Supóngase primero que T es unitario y, por lo tanto, normal. Entonces si $T(x) = \lambda x$, tenemos que $|\lambda| \cdot ||x|| = ||\lambda x|| = ||T(x)|| = ||x||$ y, por lo tanto, $|\lambda| = 1$ si $x \neq 0$.

Ahora supóngase que $|\lambda| = 1$ para todo eigenvalor λ de T, y sea $T = \lambda_1 T_1 + \ldots + \lambda_k T_k$ la descomposición espectral de T. Entonces, por el inciso (c) del Teorema 7.27,

$$\begin{aligned} TT^* &= (\lambda_1 T_1 + \ldots + \lambda_k T_k)(\bar{\lambda}_1 T_1 + \ldots + \bar{\lambda}_k T_k) \\ &= |\lambda_1|^2 T_1 + \ldots + |\lambda_k|^2 T_k \\ &= T_1 + \ldots + T_k \\ &= I. \end{aligned}$$

Por lo tanto, T es unitario. ■

Corolario 3. *Si $F = C$ y T es normal, entonces T es autoadjunto si y sólo si todo eigenvalor de T es real.*

DEMOSTRACIÓN. Sea $T = \lambda_1 T_1 + \ldots + \lambda_k T_k$ la descomposición espectral de T. Supóngase que todo eigenvalor de T es real. Entonces $T^* = \bar{\lambda}_1 T_1 + \ldots + \bar{\lambda}_k T_k = \lambda_1 T_1 + \ldots + \lambda_k T_k = T$.

La recíproca ya ha sido demostrada en el Corolario 2 del Teorema 7.15. ■

Corolario 4. *Sea T como en el teorema espectral con una descomposición espectral $T = \lambda_1 T_1 + \ldots + \lambda_k T_k$. Entonces, cada T_1 es un polinomio en T.*

DEMOSTRACIÓN. Selecciónese un polinomio $g_j (1 \leq j \leq k)$ tal que $g_j(\lambda_i) = \delta_{ij}$. Entonces, $g_j(T) = g_j(\lambda_i)T_1 + \ldots + g_j(\lambda_k)T_k = \delta_{1j}T_1 + \ldots + \delta_{kj}T_k = T_j$. ■

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas. Supóngase que los espacios subyacentes con producto interior son dimensionalmente finitos.
   (a) Todas las proyecciones son autoadjuntas.
   (b) Las proyecciones ortogonales están determinadas de manera única por su rango.
   (c) Todo operador autoadjunto es una combinación lineal de proyecciones ortogonales.

(d) Si un operador posee una descomposición espectral, también la posee su adjunto.
(e) Si T es una proyección en W, entonces $T(x)$ es el vector de W más cercano a x.
(f) Toda proyección ortogonal es un operador unitario.

2. Sean $V = R^2$, $W = L(\{(1, 2)\})$ y β la base ordenada estándar para V. Calcular $[T]_\beta$, donde T es la proyección ortogonal sobre W. Realizar lo mismo para $V = R^3$ y $W = L(\{(1, 0, 1)\})$.

3. Para cada una de las matrices A del Ejercicio 2 de la Sección 7.7

(i) Demostrar que L_A posee una descomposición espectral.
(ii) De una manera explícita, defínanse cada una de las proyecciones ortogonales en los eigenespacios de L_A.
(iii) Verifíquense los resultados utilizando el teorema espectral.

4. Sea W un subespacio dimensionalmente finito de un espacio con producto interior V. Demostrar que si T es la proyección ortogonal sobre W, entonces $I - T$ es la proyección ortogonal sobre $W^\perp$.

5. Sea V un espacio dimensionalmente finito con producto interior, y sea $T: V \to V$ una proyección.

(a) Si T es una proyección ortogonal, demostrar que $||T(x)|| \leq ||x||$ para toda $x \in V$. Dar un ejemplo de una proyección T para la cual no se cumpla esta desigualdad. Si se da la igualdad, ¿qué puede concluirse sobre T?
(b) Si T es también normal y V es compleja, demostrar que T debe ser una proyección ortogonal.

6. Si T y U son proyecciones ortogonales en un espacio con producto interior tales que $TU = T_0 = UT$, demostrar que $R(T) = R(U)^\perp$.

7. Sea T un operador normal en un espacio complejo con producto interior dimensionalmente finito V. Utilizar la descomposición espectral $\lambda_1 T_1 + \ldots + \lambda_k T_k$ de T para demostrar los siguientes incisos.

(a) Si g es un polinomio, entonces

$$g(T) = \sum_{i=1}^{k} g(\lambda_i) T_i.$$

(b) Si $T^n = T_0$ para alguna n, *entonces* $T = T_0$.
(c) $U: V \to V$ conmuta con T si y sólo si U conmuta con cada T_i.
(d) Si $U: V \to V$ es normal y conmuta con T, entonces $U = \mu_1 T_1 + \ldots + \mu_r T_r$, donde $\mu_1, \ldots, \mu_r$ son los eigenvalores (no necesariamen-

te distintos) de U. *Sugerencia:* Demostrar que los eigenespacios de T son invariantes bajo U.

(e) Existe un operador normal U en V tal que $U^2 = T$.

(f) T es invertible si $\lambda_i \neq 0$ para $1 \leq i \leq k$.

(g) T es una proyección si y sólo si todo eigenvalor de T es 1 o 0.

(h) $T = -T^*$ (tal T se denomina *antisimétrico)* si y sólo si todo λ_i es un número imaginario.

8. Utilizar el Corolario 1 del teorema espectral para demostrar que si T es un operador normal en un espacio complejo dimensionalmente finito con producto interior y U conmuta con T, entonces U conmuta con T*.

9. Refiriéndose al Ejercicio 19 de la Sección 7.7, demostrar los siguientes hechos sobre U.

(a) U*U es una proyección ortogonal sobre W.

(b) UU*U = U.

10. *Diagonalización simultánea.* Sea V un espacio complejo dimensionalmente finito con producto interior, y sean U, T: $V \rightarrow V$ operadores normales tales que TU = UT. Demostrar que existe una base ortonormal para V formada por vectores que son eigenvectores tanto de T como de U. *Sugerencia:* Emplear la sugerencia del Ejercicio 13 de la Sección 7.5 junto con el Ejercicio 8.

11. Demostrar el inciso (c) del teorema espectral.

7.10* APROXIMACION POR MINIMOS CUADRADOS

Considérese el siguiente problema: Un investigador recolecta información mediante la realización de mediciones $y_1, y_2, \ldots, y_m$ en los instantes $t_1, t_2, \ldots, t_m$, respectivamente. Por ejemplo, puede realizar mediciones sobre el desempleo en distintas fechas durante un período. Supóngase que grafica los datos $(t_1, y_1), \ldots, (t_m, y_m)$ como puntos del plano. (Véase Fig. 7.8.) A causa de la distribución de tales puntos, él piensa que existe una correlación lineal entre y y t, tal como $y = ct + d$. El investigador desearía encontrar el valor de los parámetros c y d de tal manera que la recta $y = ct + d$ represente el mejor "ajuste" posible para los datos recopilados. Una estimación del ajuste es calcular el error E que representa la suma de los cuadrados de las distancias verticales de los puntos a la recta; esto es,

$$E = \sum_{i=1}^{m} (y_i - ct_i - d)^2.$$

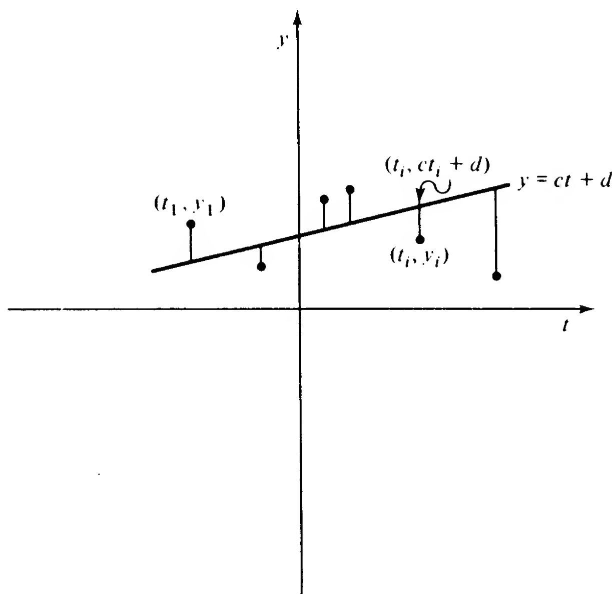


figura 7.8

Así, su problema es encontrar las constantes c y d que minimicen a E. (Para esta razón, la recta $y = ct + d$ se denomina la *recta de los mínimos cuadrados.)* Esto lo conduce a considerar el siguiente sistema de ecuaciones:

$$\begin{cases} t_1 c + d = y_1 \\ t_2 c + d = y_2 \\ \quad\cdot \\ \quad\cdot \\ \quad\cdot \\ t_m c + d = y_m, \end{cases}$$

o bien $AX = y$, donde

$$A = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ t_m & 1 \end{pmatrix}, \quad X = \begin{pmatrix} c \\ d \end{pmatrix} \quad \text{y} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_m \end{pmatrix}.$$

Nótese que $E = \|y - AX\|^2$.

Por supuesto, sería irreal suponer que tal sistema tiene una solución puesto que, en la práctica, el número de ecuaciones excede con mucho al

número de incógnitas. Pero desarrollaremos ahora un método general utilizando la teoría de las proyecciones ortogonales para encontrar un vector explícito $x_0 \in \mathsf{F}^n$ que minimice a E; esto es, dada una matriz A de $m \times n$, encontraremos a $x_0 \in \mathsf{F}^n$ tal que $||y - Ax_0|| \leq ||y - Ax||$ para todos los vectores $x \in \mathsf{F}^n$. Este método no sólo nos permitirá encontrar la función lineal que se ajuste mejor a los datos, sino también el polinomio de cualquier grado fijo que se ajuste mejor a los datos.

Requeriremos primero de algo de notación y de dos lemas sencillos.

Para x, $y \in \mathsf{F}^n$, denotaremos por $(x, y)_n$ al producto interior ordinario de x y y en F^n.

Lema 1. *Sean* A *una matriz de* m × n *sobre* F, $x \in \mathsf{F}^n$ *y* $y \in \mathsf{F}^m$. *Entonces*

$$(Ax, y)_m = (x, A^*y)_n.$$

DEMOSTRACIÓN. Demostraremos el resultado para x y y contenidas en las bases ordenadas estándar, respectivamente, para F^n y F^m. Dejamos el caso general para el lector. Sean $x = e_i$ y $y = e'_j$ dichos elementos. Utilizando el Teorema 2.15, tenemos

$$(Ae_i, e'_j)_m = (A^i, e'_j)_m = A_{ji} \quad \text{y} \quad (e_i, A^*e'_j)_n = (e_i, (A^*)^j)_n = \overline{(A^*)_{ij}} = A_{ji},$$

donde A^i y $(A^*)^j$ son las columnas i de A y j de A^*, respectivamente. ■

Lema 2. *Sea* A *una matriz de* m × n *sobre* F. *Entonces rango*(A^*A) = *rango*(A).

DEMOSTRACIÓN. Sólo tenemos que demostrar que, para $x \in \mathsf{F}^n$, $A^*Ax = 0$ si y sólo si $Ax = 0$. Claramente $Ax = 0$ implica que $A^*Ax = 0$. Por ello, supóngase que $A^*Ax = 0$. Entonces $0 = (A^*Ax, x)_n = (Ax, A^{**}x)_m = (Ax, Ax)_m$, de modo que $Ax = 0$. ■

Corolario. *Si* A *es una matriz de* m × n *tal que rango*(A) = n *(esto es,* A *tiene "rango completo"), entonces* A^*A *es invertible.*

Ahora considérese el sistema $AX = y$, donde A es una matriz de $m \times n$ y $y \in \mathsf{F}^m$. Defínase a $\mathsf{W} = \{Ax\colon x \in \mathsf{F}^n\}$, esto es, $\mathsf{W} = \mathsf{R}(\mathsf{L}_A)$. Haciendo a T la proyección ortogonal sobre W, escójase a $x_0 \in \mathsf{F}^n$ tal que $\mathsf{T}(y) = Ax_0$. Entonces, por el Teorema 7.25, $||\mathsf{T}(y) - y|| \leq ||u - y||$ para toda $u \in \mathsf{W}$; esto es, $||Ax_0 - y|| \leq ||Ax - y||$ para toda $x \in \mathsf{F}^n$.

Para desarrollar un método práctico para encontrar tal x_0, observamos que, como T es una proyección ortogonal, $Ax_0 - y = \mathsf{T}(y) - y \in \mathsf{W}_1$ y entonces $(Ax, Ax_0 - y)_m = 0$ para toda $x \in \mathsf{F}^n$. Luego, por el Lema 1, tenemos que $(x, A^*(Ax_0 - y))_n = 0$ para toda $x \in \mathsf{F}^n$; esto es, $A^*(Ax_0 - y) = 0$. Así, únicamente tenemos que encontrar una solución para $A^*AX = A^*y$. Si, además, suponemos que rango$(A) = n$, entonces por el Lema 2 tenemos que $x_0 = (A^*A)^{-1}A^*y$. Podemos resumir esta exposición en el teorema siguiente.

Teorema 7.28. *Sea* $A \in M_{m\times n}(F)$ *y* $y \in F^m$. *Entonces, existe* $x_0 \in F^n$ *tal que* $(A^*A)x_0 = A^*y$ *y* $\|Ax_0 - y\| \leq \|Ax - y\|$ *para toda* $x \in F^n$. *Además, si* *rango*$(A) = n$, *entonces* $x_0 = (A^*A)^{-1}A^*y$.

Volviendo con nuestro investigador, supongamos que éste recopila los siguientes datos: (1, 2), (2, 3), (3, 5) y (4, 7). Entonces

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} \quad \text{y} \quad y = \begin{pmatrix} 2 \\ 3 \\ 5 \\ 7 \end{pmatrix};$$

por lo tanto

$$A^*A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 30 & 10 \\ 10 & 4 \end{pmatrix},$$

y entonces

$$(A^*A)^{-1} = \frac{1}{20}\begin{pmatrix} 4 & -10 \\ -10 & 30 \end{pmatrix}.$$

Por consiguiente

$$x_0 = \begin{pmatrix} c \\ d \end{pmatrix} = \frac{1}{20}\begin{pmatrix} 4 & -10 \\ -10 & 30 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \\ 5 \\ 7 \end{pmatrix} = \begin{pmatrix} 1.7 \\ 0 \end{pmatrix}.$$

Así, la recta $y = 1.7t$ es la recta de mínimos cuadrados. El error E puede calcularse directamente como $\|Ax_0 - y\|^2 = 0.3$.

El método anterior puede también ser aplicado si el investigador desea ajustar una parábola $y = ct^2 + dt + e$ a los datos. En este caso, él utilizaría

$$\begin{pmatrix} t_1^2 & t_1 & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ t_m^2 & t_m & 1 \end{pmatrix}$$

como la matriz A.

Finalmente, supóngase en el caso lineal que el investigador escogió sus instantes, t_i, para satisfacer la expresión

$$\sum_{i=1}^{m} t_i = 0.$$

Entonces, las dos columnas de A serían ortogonales y A^*A sería una matriz diagonal. (Véase el Ejercicio 1.) Esto, por supuesto, simplificaría mucho los cálculos.

Soluciones mínimas

En la exposición anterior mostramos que, si rango$(A) = n$, entonces existe un elemento $x_0 \in F^n$ único tal que Ax_0 es el punto en W más cercano a y. Por supuesto, si rango$(A) < n$, existirá un número infinito de estos vectores. Es a menudo deseable encontrar un vector tal que su norma sea mínima. Para lo siguiente, haremos que $b = Ax_0$; esto es, $b = T(y)$, donde T es la proyección ortogonal sobre W. Entonces, el sistema $AX = b$ tiene al menos una solución. Una solución s se llama *solución mínima* si $\|s\| \leq \|u\|$ para todas las otras soluciones u de $AX = b$.

Teorema 7.29. *Sean* $A \in M_{m\times n}(F)$ *y* $b \in F^m$. *Supóngase que* $AX = b$ *tiene al menos una solución. Entonces*

(a) *Existe exactamente una solución mínima* s *de* $AX = b$ *y* $s \in R(L_{A^*})$.

(b) s *es la única solución de* $AX = b$ *ubicada en* $R(L_{A^*})$; *esto es, si* u *es una solución de* $(AA^*)X = b$, *entonces* $s = A^*u$.

DEMOSTRACIÓN. Para simplicidad en la notación, escribiremos $N(A) = N(L_A)$ y $R(A^*) = R(L_{A^*})$. Por el Teorema 7.6 y el Ejercicio 12 de la Sección 7.3, tenemos que $F^n = N(A)^\perp \oplus N(A) = R(A^*) \oplus N(A)$. Sea x una solución cualquiera de $AX = b$. Entonces, por lo anterior, $x = s + y$, donde $s \in R(A^*)$ y $y \in N(A)$. Nótese que $b = Ax = As + Ay = As$, de manera que s es una solución de $AX = b$ que está en $R(A^*)$. Para demostrar (a), necesitamos únicamente demostrar que s es la única solución mínima. Sea v cualquier solución de $AX = b$. Por el Teorema 3.8, tenemos que $v = s + u$, donde $u \in N(A)$. Como $s \in R(A^*) = N(A)^\perp$, tenemos, de acuerdo con el Ejercicio 10 de la Sección 7.1, que $\|v\|^2 = \|s + u\|^2 = \|s\|^2 + \|u\|^2 \geq \|s\|^2$. Luego, s es una solución mínima. Podemos ver igualmente del cálculo anterior que si $\|v\| = \|s\|$, entonces $u = 0$ y $v = s$. Por lo tanto, s es la única solución mínima de $AX = b$.

Con el objeto de demostrar el inciso (b) supondremos que v es también una solución de $AX = b$ que está en $R(A^*)$. Entonces $v - s \in R(A^*) \cap N(A) = \{0\}$, y así $v = s$. ■

Ejemplo 32. Considérese el sistema

$$\begin{cases} x + 2y + z = 4 \\ x - y + 2z = -11 \\ x + 5y = 19. \end{cases}$$

Sean

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & -1 & 2 \\ 1 & 5 & 0 \end{pmatrix} \quad \text{y} \quad b = \begin{pmatrix} 4 \\ -11 \\ 19 \end{pmatrix}.$$

Para encontrar la solución mínima del sistema, debemos encontrar una solución de $AA^*X = b$. Ahora bien,

$$AA^* = \begin{pmatrix} 6 & 1 & 11 \\ 1 & 6 & -4 \\ 11 & -4 & 26 \end{pmatrix},$$

por lo que consideraremos el sistema

$$\begin{cases} 6x + \ y + 11z = \ \ \ 4 \\ \ x + 6y - \ 4z = -11 \\ 11x - 4y + 26z = \ \ 19, \end{cases}$$

para el cual una solución es

$$u = \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}.$$

(Cualquier solución es suficiente.) Por lo tanto,

$$s = A^*u = \begin{pmatrix} -1 \\ 4 \\ -3 \end{pmatrix}$$

es la solución mínima del sistema dado.

EJERCICIOS

1. Demostrar que si A es una matriz de $m \times n$ cuyas columnas son ortogonales, entonces A^*A es una matriz diagonal.

2. Dados los datos $(-3, 9)$, $(-2, 6)$, $(0, 2)$ y $(1, 1)$, encontrar la parábola que proporcione el ajuste de mínimos cuadrados. Calcular E.

3. Calcular la solución mínima de

$$\begin{cases} x + 2y - z = 1 \\ 2x + 3y + z = 2 \\ 4x + 7y - z = 4. \end{cases}$$

4. Sea A una matriz de $m \times n$. Demostrar que $(Ax, y)_m = (x, A^*y)_n$ para $x \in \mathsf{F}^n$ e $y \in \mathsf{F}^m$, completando así la demostración del Lema 1.

5. Para la recta de mínimos cuadrados $y = ct + d$ correspondiente a las m observaciones $(t_1, y_1), \ldots, (t_m, y_m)$, utilizar el Teorema 7.28 para derivar las *ecuaciones normales:*

$$\left(\sum_{i=1}^{m} t_i^2\right) c + \left(\sum_{i=1}^{m} t_i\right) d = \sum_{i=1}^{m} t_i y_i$$

y

$$\left(\sum_{i=1}^{m} t_i\right) c + md = \sum_{i=1}^{m} y_i.$$

Estas ecuaciones también se pueden obtener haciendo cada una de las derivadas parciales del error E iguales a cero.

7.11* FORMAS BILINEALES Y CUADRATICAS

Existe una cierta clase de funciones de valor escalar de dos variables definidas en un espacio vectorial que a menudo se considera en el estudio de temas tan diversos como la geometría y el cálculo en varias variables. Esta es la clase de "formas bilineales". Estudiaremos las propiedades básicas de esta clase dando importancia especial a las formas bilineales simétricas y consideraremos algunas de sus aplicaciones a las superficies cuadráticas y al cálculo en varias variables.

A lo largo de esta sección todas las bases serán consideradas como bases ordenadas.

Definición. *Sea* V *un espacio vectorial sobre un campo* F. *Una función* H *que va del conjunto* V × V *de pares ordenados de vectores en* V *a* F, *se llama* forma bilineal *en* V *si* H *es lineal en cada variable cuando la otra variable se mantiene fija, esto es, si*

(a) $H(ax_1 + x_2, y) = aH(x_1, y) + H(x_2, y)$ *para toda* x_1, x_2, $y \in V$ *y* $a \in F$.

(b) $H(x, ay_1 + y_2) = aH(x, y_1) + H(x, y_2)$ *para toda* x, y_1, $y_2 \in V$ *y* $a \in F$.

Representaremos al conjunto de formas bilineales en V mediante $\mathcal{B}(V)$. Obsérvese que un producto interior en un espacio real V es una forma bilineal.

Ejemplo 33. Defínase una función H: $R^2 \times R^2 \to R$ mediante

$$H\left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}\right) = 2a_1b_1 + 3a_1b_2 + 4a_2b_1 - a_2b_2 \qquad \text{para } \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \in R^2.$$

Podemos verificar directamente que H es una forma bilineal en R^2. Será más ilustrativo y menos tedioso, sin embargo, observar que si

$$A = \begin{pmatrix} 2 & 3 \\ 4 & -1 \end{pmatrix}, \qquad x = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad \text{y} \quad y = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix},$$

entonces

$$H(x, y) = x^t A y.$$

La bilinealidad de H se infiere ahora directamente de la propiedad distributiva de la multiplicación de matrices sobre la suma de matrices.

La forma bilineal anterior es un caso especial de la siguiente situación que es más general.

Ejemplo 34. Sea $V = F^n$ el espacio vectorial de todos los vectores columna de longitud n sobre un campo F. Para cualquier matriz A de $n \times n$ con elementos de F, defínase a H: $V \times V \to F$ mediante

$$H(x, y) = x^t A y \qquad \text{para } x, y \in V.$$

Nótese que, como x y y son matrices de $n \times 1$ y A es una matriz de $n \times n$, $H(x, y)$ es una matriz de 1×1 para toda $x, y \in V$. Identificamos a esta matriz con su elemento único. Como en el Ejemplo 33, la bilinealidad de H se obtiene de la propiedad distributiva de la multiplicación sobre la suma de matrices. Por ejemplo, si $a \in F$ y $x_1, x_2, y \in V$, entonces

$$\begin{aligned} H(ax_1 + x_2, y) &= (ax_1 + x_2)^t A y = (ax_1^t + x_2^t) A y \\ &= ax_1^t A y + x_2^t A y \\ &= aH(x_1, y) + H(x_2, y). \end{aligned}$$

Enumeraremos ahora algunas propiedades que tienen todas las formas bilineales. Las demostraciones se dejan al lector. (Véase Ejercicio 2.)

Para cualquier forma bilineal H en un espacio vectorial V sobre un campo F:

1. Si para cualquier $x \in V$ las funciones L_x, R_x: $V \to F$ se definen mediante $L_x(y) = H(x, y)$ y $R_x(y) = H(y, x)$ para toda $y \in V$, entonces L_x y R_x son lineales.
2. $H(0, x) = H(x, 0) = 0$ para toda $x \in V$.
3. Si $x, y, z, w \in V$, entonces

$$H(x + y, z + w) = H(x, z) + H(x, w) + H(y, z) + H(y, w).$$

4. Si J: $V \times V \to F$ está definida mediante $J(x, y) = H(y, x)$, entonces J es una forma bilineal.

Para un espacio vectorial V, $H_1, H_2 \in \mathcal{B}(V)$ y cualquier escalar a, definamos la *suma* $H_1 + H_2$ y el *producto* aH_1 mediante las ecuaciones

$$(H_1 + H_2)(x, y) = H_1(x, y) + H_2(x, y)$$

y

$$(aH_1)(x, y) = a(H_1(x, y)) \qquad \text{para toda } x, y \in V.$$

Es un ejercicio sencillo verificar que $H_1 + H_2$ y aH_1 son de nuevo formas bilineales. No es de sorprenderse que $\mathcal{B}(V)$ sea un espacio vectorial con respecto a estas operaciones.

Teorema 7.30. *Para cualquier espacio vectorial* V, $\mathcal{B}(V)$ *es un espacio vectorial con respecto a las anteriores definiciones de suma y de producto.*

DEMOSTRACIÓN. Ejercicio.

Sea V un espacio vectorial n-dimensional con una base $\beta = \{x_1, x_2, \ldots, x_n\}$. Para cualquier forma bilineal $H \in \mathcal{B}(V)$ podemos asociar con H una matriz A de $n \times n$ cuyo elemento del renglón i y columna j esté definido mediante

$$A_{ij} = H(x_i, x_j) \qquad \text{para toda } i, j = 1, 2, \ldots, n.$$

Definición. *La matriz anterior* A *será llamada la* representación matricial de H con respecto a la base β.

Podemos, por lo tanto, definir un mapeo ψ_β de $\mathcal{B}(V)$ en $M_{n\times n}(F)$, donde F es el campo de escalares para V, tal que para cualquier $H \in \mathcal{B}(V)$, $\psi_\beta(H) = A$, donde A es la representación matricial H con respecto a β.

Ejemplo 35. Considérese la forma bilineal de H del Ejemplo 33. Sean

$$\beta = \left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}\right\} \quad \text{y} \quad B = \psi_\beta(H).$$

Entonces

$$B_{11} = H\left(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}\right) = 2 + 3 + 4 - 1 = 8,$$

$$B_{12} = H\left(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}\right) = 2 - 3 + 4 + 1 = 4,$$

$$B_{21} = H\left(\begin{pmatrix}1\\-1\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}\right) = 2 + 3 - 4 + 1 = 2,$$

$$B_{22} = H\left(\begin{pmatrix}1\\-1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}\right) = 2 - 3 - 4 - 1 = -6.$$

Así,

$$\psi_\beta(H) = \begin{pmatrix}8 & 4\\2 & -6\end{pmatrix}.$$

Si γ es la base estándar para R^2, el lector podrá verificar que

$$\psi_\gamma(H) = \begin{pmatrix}2 & 3\\4 & -1\end{pmatrix}.$$

Teorema 7.31. *Para cualquier espacio vectorial* n*-dimensional* V *sobre un campo* F *y cualquier base* β *para* V, ψ_γ *es un isomorfismo de* $\mathcal{B}(V)$ *en* $M_{n\times n}(F)$.

DEMOSTRACIÓN. Dejaremos al lector 'a demostración de que ψ_β es una transformación lineal.

Para demostrar que ψ_β es uno-a-uno, supóngase que $H \in \mathcal{B}(V)$ y $\psi_\beta(H) = O$, la matriz nula. Deseamos demostrar que H es trivial, esto es, $H(x, y) = 0$ para toda $x, y \in V$. Fíjese una $x_i \in \beta$ y recuérdese la función L_{x_i}: $V \to F$ definida mediante $L_{x_i}(x) = H(x_i, x)$ para toda $x \in V$. De acuerdo con la propiedad 1 dada en la p. 469, L_{x_i} es lineal; por hipótesis, $L_{x_i}(x_j) = H(x_i, x_j) = 0$ para toda $x_j \in \beta$. Por lo tanto, L_{x_i} es la transformación nula de V en F. Entonces

$$H(x_i, x) = L_{x_i}(x) = 0 \qquad \text{para toda } x \in V \text{ y } x_i \in \beta. \qquad (18)$$

Fíjese luego una $y \in V$ arbitraria y recuérdese el mapeo R_γ: $V \to F$ definido mediante $R_\gamma(x) = H(x, y)$ para toda $x \in V$. De nuevo R_γ es lineal. Pero por la Ecuación (18) $R_\gamma(x_i) = H(x_i, y) = 0$ para cualquier $x_i \in \beta$. Luego, R_γ es trivial, y concluimos que $H(x, y) = R_\gamma(x) = 0$ para toda $x, y \in V$. Así, tenemos que H es trivial y ψ_β es, por lo tanto, uno-a-uno.

Para demostrar que ψ_β es sobreyectivo, sea $A \in M_{n\times n}(F)$. Recuérdese el isomorfismo ϕ_β: $V \to F^n$ definido en la Sección 2.4. Para $x \in V$ tomaremos a $\phi_\beta(x) \in F^n$ como un vector columna. Defínase un mapeo H: $V \times V: \to F$ mediante

$$H(x, y) = [\phi_\beta(x)]^t A[\phi_\beta(y)] \qquad \text{para toda } x, y \in V.$$

Por el Ejemplo 34, $H \in \mathcal{B}(V)$. Demostraremos que $\psi_\beta(H) = A$. Si $x_i, x_j \in \beta$, entonces $\phi_\beta(x_i) = e_i$ y $\phi_\beta(x_j) = e_j$. En consecuencia, para cualesquiera i y j,

$$H(x_i, x_j) = [\phi_\beta(x_i)]^t A[\phi_\beta(x_j)] = e_i^t A e_j = A_{ij}.$$

Concluimos que $\psi_\beta(H) = A$ y, por lo tanto, ψ_β es sobreyectivo. ■

Corolario 1. *Para cualquier espacio vectorial* n*-dimensional* V, $\mathcal{B}(V)$ *es de dimensión* n^2.

DEMOSTRACIÓN. Ejercicio.

El corolario siguiente queda establecido fácilmente al repasar la demostración del Teorema 7.31.

Corolario 2. *Sea* V *un espacio vectorial* n*-dimensional sobre el campo* F *con una base* β. *Si* $H \in \mathcal{B}(V)$ *y* $A \in M_{n\times n}(F)$, *entonces* $\psi_\beta(H) = A$ *si y sólo si* $H(x, y) = [\phi_\beta(x)]^t A[\phi_\beta(y)]$ *para toda* $x, y \in V$.

El siguiente corolario es ahora una consecuencia inmediata del Corolario 2.

Corolario 3. *Para cualquier campo* F, *un entero positivo* n *y* $H \in \mathcal{B}(F^n)$, *existe una matriz única* $A \in M_{n \times n}(F)$, *a saber* $A = \psi_\beta(H)$, *tal que*

$$H(x, y) = x^t A y \qquad \textit{para toda } x, y \in F^n,$$

donde β *es la base estándar para* F^n.

Parece existir una analogía entre las formas bilineales y los operadores lineales en el hecho de que ambos están asociados con una matriz cuadrada única y en que esta correspondencia depende de la selección de una base para el espacio vectorial. Tal como en el caso de los operadores, uno puede hacerse la pregunta: ¿Cómo se modifica la matriz correspondiente a una forma bilineal fija cuando se cambia la base? Como ya lo hemos visto, cuando surgió la pregunta para el caso de los operadores lineales, ésta condujo al estudio de la relación de similitud en las matrices cuadradas. En el caso de las formas bilineales nos orientaremos al estudio de otra relación en las matrices cuadradas, la relación de "congruencia".

Definición. *Se dice que dos matrices* $A, B \in M_{n \times n}(F)$ *son* congruentes *si existe una matriz invertible* $Q \in M_{n \times n}(F)$ *tal que*

$$Q^t A Q = B.$$

Se puede ver fácilmente que la congruencia es una relación de equivalencia. (Véase el Ejercicio 11.)

El teorema siguiente relaciona la congruencia con la representación matricial de una forma bilineal.

Teorema 7.32. *Sea* V *un espacio vectorial dimensionalmente finito con bases* $\beta = \{x_1, x_2, \ldots, x_n\}$ *y* $\gamma = \{y_1, y_2, \ldots, y_n\}$ *y sea* Q *la matriz de cambio de coordenadas que transforma las coordenadas de* γ *en coordenadas de* β. *Entonces, para cualquier* $H \in \mathcal{B}(V)$, $\psi_\gamma(H) = Q^t \psi_\beta(H) Q$. *En particular,* $\psi_\gamma(H)$ *y* $\psi_\beta(H)$ *son congruentes.*

DEMOSTRACIÓN. Existen fundamentalmente dos demostraciones de este teorema. La primera implica un cálculo directo, mientras que la otra se obtiene inmediatamente de una observación acuciosa. Presentaremos la primera demostración y dejaremos la última como ejercicio. (Véase el Ejercicio 12.)

Supóngase que $A = \psi_\beta(H)$ y $B = \psi_\gamma(H)$. Entonces, para cualesquiera i y j tales que $1 \leq i, j \leq n$,

$$y_i = \sum_{k=1}^{n} Q_{ki} x_k \quad \text{y} \quad y_j = \sum_{r=1}^{n} Q_{rj} x_r.$$

Así,

$$\begin{aligned} B_{ij} = H(y_i, y_j) &= H\left(\sum_{k=1}^{n} Q_{ki}x_k, y_j\right) \\ &= \sum_{k=1}^{n} Q_{ki}H(x_k, y_j) \\ &= \sum_{k=1}^{n} Q_{ki}H\left(x_k, \sum_{r=1}^{n} Q_{rj}x_r\right) \\ &= \sum_{k=1}^{n} Q_{ki} \sum_{r=1}^{n} Q_{rj}H(x_k, x_r) \\ &= \sum_{k=1}^{n} Q_{ki} \sum_{r=1}^{n} Q_{rj}A_{kr} \\ &= \sum_{k=1}^{n} Q_{ki} \sum_{r=1}^{n} A_{kr}Q_{rj} \\ &= \sum_{k=1}^{n} Q_{ki}(AQ)_{kj} \\ &= \sum_{k=1}^{n} Q^t_{ik}(AQ)_{kj} = (Q^tAQ)_{ij}. \end{aligned}$$

Por lo tanto, $B = Q^tAQ$. ■

El corolario siguiente es el recíproco del Teorema 7.32.

Corolario. *Sea* A, B $\in$ M$_{n\times n}$(F). *Si* A *y* B *son congruentes, entonces existen un espacio vectorial* n*-dimensional* V *sobre* F, *bases* β *y* γ *para* V, *y una forma bilineal* H *en* V, *tales que*

$$\psi_\beta(\mathrm{H}) = \mathrm{A} \quad y \quad \psi_\gamma(\mathrm{H}) = \mathrm{B}.$$

DEMOSTRACIÓN. Supóngase que Q es una matriz invertible para la que $B = Q^tAQ$. Sean $\mathsf{V} = \mathsf{F}^n$, $\beta = \{e_1, e_2, \ldots, e_n\}$ la base estándar para F^n y H la preimagen de A bajo ψ_β. Sea $\gamma = \{Q^1, Q^2, \ldots, Q^n\}$ el conjunto de columnas de Q. Entonces, γ es una base para F^n y Q es la matriz de cambio de coordenadas que transforma las coordenadas de γ en coordenadas de β. Luego, por el Teorema 7.32, $B = Q^tAQ = Q^t\psi_\beta(H)Q = \psi_\gamma(H)$. ■

Tal como el problema de diagonalización para operadores lineales, existe un problema de "diagonalización" semejante para formas bilineales, a saber, el problema de la determinación de aquellas formas bilineales para las que existen representaciones matriciales diagonales. Como veremos, las formas bilineales "diagonalizables" son aquellas que son "simétricas".

Definición. *Una forma bilineal* H *sobre un espacio vectorial* V *se llama* simétrica, *si* H(x, y) = H(y, x) *para toda* x, y $\in$ V.

Como el nombre lo sugiere, las formas bilineales simétricas corresponden a matrices simétricas.

Teorema 7.33. *Sea* V *un espacio vectorial dimensionalmente finito. Para* H ∈ ℬ(V) *las siguientes proposiciones son equivalentes:*

(a) H *es simétrica.*

(b) *Para cualquier base* γ *para* V, $\psi_\gamma(\mathrm{H})$ *es una matriz simétrica.*

(c) *Existe una base* β *para* V *tal que* $\psi_\beta(\mathrm{H})$ *es una matriz simétrica.*

DEMOSTRACIÓN. Primero demostraremos que (a) implica a (b). Supóngase que H es simétrica. Sea $\gamma = \{y_1, y_2, \ldots, y_n\}$ una base para V, y sea $B = \psi_\gamma(H)$. Entonces, para cualquier i y cualquier j, $B_{ij} = H(y_i, y_j) = H(y_j, y_i) = B_{ji}$. Luego, B es una matriz simétrica, demostrando así a (b).

Claramente (b) implica a (c).

Finalmente, demostraremos que (c) implica a (a). Supóngase que, para alguna base $\beta = \{x_1, x_2, \ldots, x_n\}$, $\psi_\beta(H) = A$ es una matriz simétrica. Defínase a J: $\mathrm{V} \times \mathrm{V} \to F$, donde F es el campo de los escalares para V, mediante $J(x, y) = H(y, x)$ para toda $x, y \in \mathrm{V}$. Por la propiedad 4 dada en la p. 469, $J \in$ ℬ(V). Sea $C = \psi_\beta(J)$. Entonces, para cualquier i y cualquier j,

$$C_{ij} = J(x_i, x_j) = H(x_j, x_i) = A_{ji} = A_{ij}.$$

Por lo tanto, $C = A$. Como ψ_β es uno-a-uno, concluimos que $J = H$ y, por lo tanto, $H(y, x) = J(x, y) = H(x, y)$ para toda $x, y \in \mathrm{V}$ y entonces H es simétrica, lo que demuestra a (a). ∎

Definición. *Una forma bilineal* H *en un espacio vectorial dimensionalmente finito* V *se llama* diagonalizable *si existe una base* β *para* V *tal que* $\psi_\beta(\mathrm{H})$ *sea una matriz diagonal.*

Corolario. *Sea* V *un espacio vectorial dimensionalmente finito. Para cualquier* H ∈ ℬ(V), *si* H *es diagonalizable, entonces* H *es simétrica.*

DEMOSTRACIÓN. Supóngase que H es diagonalizable. Entonces, existe una base β para V tal que $\psi_\beta(H) = D$ es una matriz diagonal. Trivialmente, D es una matriz simétrica. Luego, de acuerdo con el Teorema 7.33, H es simétrica. ∎

Desafortunadamente, la recíproca no es cierta, como se ilustra en el ejemplo siguiente.

Ejemplo 36. Sea $F = Z_2$ (véase Apéndice C), y sea $\mathrm{V} = \mathrm{F}^2$. Defínase a H: $\mathrm{V} \times \mathrm{V} \to F$ mediante

$$H\left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}\right) = a_1 b_2 + a_2 b_1.$$

Se ve claramente que H es simétrica. De hecho, si β es la base estándar para F^2, entonces

$$A = \psi_\beta(H) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

es una matriz simétrica. Demostraremos que el considerar que H es diagonalizable conduce a una contradicción.

Supóngase que H es diagonalizable. Entonces, existe una base γ para F^2 tal que $B = \psi_\gamma(H)$ es una matriz diagonal. Luego, por el Teorema 7.32, existe una matriz invertible Q tal que $B = Q^tAQ$. Como Q es invertible, rango(B) = rango$(A) = 2$. Luego, B es una matriz diagonal cuyos elementos de la diagonal son no nulos. Como el único elemento no nulo de F es 1,

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Supóngase

$$Q = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Entonces

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = B = Q^tAQ = \begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ac + ac & bc + ad \\ bc + ad & bd + bd \end{pmatrix}.$$

Pero $p + p = 0$ para toda $p \in F$ y entonces $ac + ac = 0$. Así, comparando los elementos superiores izquierdos de las matrices de la ecuación anterior, concluimos que $1 = 0$, lo que es una contradicción. En consecuencia, H no es diagonalizable.

La forma bilineal del Ejemplo 36 es anómala. La razón de que no sea diagonalizable parte del hecho de que el campo escalar Z_2 es de característica dos. Si F no es de característica dos, entonces $1 + 1$ es invertible. Bajo estas circunstancias representaremos a "$1 + 1$" como "2" y su inverso multiplicativo como $\frac{1}{2}$.

Antes de demostrar la recíproca del corolario del Teorema 7.33 para campos escalares distintos a los de característica dos, debemos establecer el lema siguiente.

Lema. *Sea* H *una forma bilineal simétrica no trivial en un espacio vectorial* V *sobre un campo* F *de característica distinta de dos. Entonces, existe un elemento* $x \in V$ *tal que* $H(x, x) \neq 0$.

DEMOSTRACIÓN. Supóngase que para algunas $v, w \in V$, $H(v, w) \neq 0$. Si $H(v, v) \neq 0$ o $H(w, w) \neq 0$, no hay nada que demostrar. De lo contra-

rio, supóngase que $H(v, v) = H(w, w) = 0$. Haciendo $x = v + w$, tenemos que

$$\begin{aligned} H(x, x) &= H(v, v) + H(v, w) + H(w, v) + H(w, w) \\ &= 2H(v, w) \neq 0 \end{aligned}$$

puesto que $2 \neq 0$ y $H(v, w) \neq 0$. ■

Teorema 7.34. *Sea* V *un espacio vectorial dimensionalmente finito sobre un campo* F *de característica distinta de dos. Entonces, toda forma simétrica bilineal en* V *es diagonalizable.*

DEMOSTRACIÓN. Utilizaremos inducción matemática sobre $n = \dim(V)$. Si $n = 1$, todo miembro de $\mathcal{B}(V)$ es diagonalizable. Supóngase que el teorema es válido para espacios vectoriales de dimensión menor que n para algún entero fijo $n > 1$. Si H es la forma bilineal trivial, entonces, por supuesto, H es diagonalizable. Supóngase que H es simétrica y no trivial. Entonces, de acuerdo con el lema, existe un elemento $x \in V$ (necesariamente no nulo) tal que $H(x, x) \neq 0$. Defínase a L: $V \to F$ mediante $L(z) = H(x, z)$ para toda $z \in V$. Entonces, L es lineal y como $L(x) = H(x, x) \neq 0$, L es no trivial. En consecuencia, $\text{rango}(L) = 1$ y, por lo tanto, $\dim(N(L)) = n - 1$. La restricción de H a $N(L)$ es evidentemente una forma bilineal simétrica en un espacio vectorial de dimensión $n - 1$. Luego, por la hipótesis de inducción, existe una base $\{x_1, x_2, \ldots, x_{n-1}\}$ para $N(L)$ tal que $H(x_i, x_j) = 0$ para $i \neq j$ $(1 \leq i, j \leq n - 1)$. Hágase $x_n = x$. Entonces $x_n \notin N(L)$, y por lo tanto $\beta = \{x_1, \ldots, x_n\}$ es una base para V. También $H(x_i, x_n) = H(x_n, x_i) = 0$ para $i = 1, 2, \ldots, n - 1$. Concluimos que $\psi_\beta(H)$ es una matriz diagonal y entonces H es diagonalizable. ■

Corolario. *Sea* F *un campo que no tiene característica dos. Si* $A \in M_{n\times n}(F)$ *es una matriz simétrica, entonces* A *es congruente con una matriz diagonal.*

DEMOSTRACIÓN. Ejercicio.

Sea A una matriz simétrica de $n \times n$ con elementos de un campo que no tenga característica dos. Por el corolario del Teorema 7.34, A es congruente con una matriz diagonal. Mostraremos cómo encontrar una matriz diagonal D y una matriz invertible Q tales que $Q^tAQ = D$. El lector debería repasar la Sección 3.1 para recordar la relación entre las matrices elementales y las operaciones elementales con matrices.

Si E es una matriz elemental de $n \times n$, entonces AE se obtiene de A mediante una cierta operación elemental con columnas en A. De acuerdo con el Ejercicio 20, E^tA se obtiene a partir de A por medio de la misma operación realizada en los renglones en vez de en las columnas de A. Entonces, E^tAE se obtiene a partir de A realizando una elemental operación en las columnas de A, y luego realizando la misma operación

en los renglones de la matriz AE. (Nótese que puede invertirse el orden de las operaciones.) Ahora supóngase que Q es una matriz invertible y que D es una matriz diagonal tal que $Q^tAQ = D$. Por el Corolario 3 del Teorema 3.5, Q es un producto de matrices elementales, $Q = E_1E_2 \dots$ $\dots E_k$. Entonces, $D = Q^tAQ = E_k^tE_{k-1}^t \dots E_1^tAE_1E_2 \dots E_k$.

Sobre la base de la ecuación anterior, concluimos que por medio de algunas operaciones elementales con columnas y las operaciones correspondientes con renglones, A puede ser transformada en una matriz diagonal D. Además, si $E_1, E_2, \dots, E_k$ son las matrices elementales correspondientes a las operaciones elementales con columnas (indexadas en el orden en que han de ser realizadas) y si $Q = E_1E_2 \dots E_k$, entonces $Q^tAQ = D$.

El enunciado anterior proporciona la clave para encontrar D y Q para una A dada.

Ejemplo 37. Supóngase que

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}.$$

Empezamos utilizando operaciones elementales con columnas para insertar un cero en el primer renglón, segunda columna; en este caso debemos restar dos veces la primera columna de A de la segunda columna de A. La operación correspondiente con renglones implicaría restar dos veces el primer renglón del segundo renglón. Sea E_1 la matriz elemental correspondiente a la anterior operación elemental con columnas. Entonces,

$$E_1 = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix};$$

en consecuencia,

$$AE_1 = \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix} \quad \text{y} \quad E_1^tAE = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Obsérvese que como E_1 produjo un cero en el renglón 1, columna 2, E_1^t produjo un cero en el renglón 2, columna 1. Luego, para

$$Q = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \quad \text{y} \quad D = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad Q^tAQ = D.$$

Considérese ahora

$$A = \begin{pmatrix} 0 & 1 & 2 \\ 1 & -1 & 3 \\ 2 & 3 & 4 \end{pmatrix}.$$

Es conveniente tener ± 1 en la posición renglón 1, columna 1, y utilizarlo para eliminar a todos los demás elementos del primer renglón y de la primera columna de A. Entonces, principiemos intercambiando la pri-

mera y la segunda columnas de A. La matriz elemental que corresponde a esta operación con columnas es

$$E_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Claramente

$$E_1^t A E_1 = \begin{pmatrix} -1 & 1 & 3 \\ 1 & 0 & 2 \\ 3 & 2 & 4 \end{pmatrix}.$$

Esta matriz se obtiene intercambiando las dos primeras columnas de A para obtener AE y luego intercambiando los dos primeros renglones de AE. Luego, produzcamos un cero en el primer renglón, segunda columna y en el segundo renglón, primera columna de $E_1^t AE_1$ sumando la primera columna de $E_1^t AE_1$ a la segunda columna de $E_1^t AE_1$ y continuando esta operación con su correspondiente operación con renglones. Finalmente, sumemos tres veces la primera columna a la tercera columna y prosigamos con la operación correspondiente con renglones. Nótese que las operaciones con columnas pueden realizarse en sucesión antes de realizar las operaciones con renglones. Entonces, si

$$E_2 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{y} \quad E_3 = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$E_3^t E_2^t (E_1^t A E_1) E_2 E_3 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 5 \\ 0 & 5 & 13 \end{pmatrix}.$$

El lector puede ver ahora fácilmente que haciendo

$$E_4 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -5 \\ 0 & 0 & 1 \end{pmatrix},$$

tenemos

$$E_4^t (E_3^t E_2^t E_1^t A E_1 E_2 E_3) E_4 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -12 \end{pmatrix}.$$

Entonces con $Q = E_1 E_2 E_3 E_4$ y

$$D = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -12 \end{pmatrix},$$

$Q^t AQ = D$.

El lector deberá justificar el siguiente método (semejante al introducido en la Sección 3.2 para calcular la inversa de una matriz) para calcular Q^t (y por lo tanto Q) sin registrar por separado a cada una de las matrices elementales: utilizar una secuencia de operaciones elementales con columnas seguidas de las correspondientes operaciones elementales con renglones para modificar la matriz aumentada $(A|I)$ en la forma $(D|B)$, donde D es una matriz diagonal. Entonces $B = Q^t$.

En el ejemplo anterior este método debería producir la siguiente secuencia de matrices:

$$(A|I) = \left(\begin{array}{ccc|ccc} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & -1 & 3 & 0 & 1 & 0 \\ 2 & 3 & 4 & 0 & 0 & 1 \end{array}\right), \quad \left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ -1 & 1 & 3 & 0 & 1 & 0 \\ 3 & 2 & 4 & 0 & 0 & 1 \end{array}\right),$$

$$\left(\begin{array}{ccc|ccc} -1 & 1 & 3 & 0 & 1 & 0 \\ 1 & 0 & 2 & 1 & 0 & 0 \\ 3 & 2 & 4 & 0 & 0 & 1 \end{array}\right), \quad \left(\begin{array}{ccc|ccc} -1 & 0 & 3 & 0 & 1 & 0 \\ 1 & 1 & 2 & 1 & 0 & 0 \\ 3 & 5 & 4 & 0 & 0 & 1 \end{array}\right),$$

$$\left(\begin{array}{ccc|ccc} -1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 5 & 1 & 1 & 0 \\ 3 & 5 & 4 & 0 & 0 & 1 \end{array}\right), \quad \left(\begin{array}{ccc|ccc} -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 5 & 1 & 1 & 0 \\ 3 & 5 & 13 & 0 & 0 & 1 \end{array}\right),$$

$$\left(\begin{array}{ccc|ccc} -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 5 & 1 & 1 & 0 \\ 0 & 5 & 13 & 0 & 3 & 1 \end{array}\right), \quad \left(\begin{array}{ccc|ccc} -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 5 & -12 & 0 & 3 & 1 \end{array}\right),$$

y

$$\left(\begin{array}{ccc|ccc} -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & -12 & -5 & -2 & 1 \end{array}\right) = (D|Q^t).$$

Por lo tanto,

$$D = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -12 \end{pmatrix}, \quad Q^t = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ -5 & -2 & 1 \end{pmatrix} \text{ y } Q = \begin{pmatrix} 0 & 1 & -5 \\ 1 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Formas cuadráticas

Asociadas con las formas bilineales simétricas existen funciones llamadas "formas cuadráticas".

Definición. *Sea* V *un espacio vectorial sobre un campo* F. *Una función* K: V → F *se llama* forma cuadrática *si existe una forma bilineal simétrica* H ∈ $\mathcal{B}$(V) *tal que*

$$K(x) = H(x, x) \text{ para toda } x \in V. \tag{19}$$

Si el campo F no es de característica dos, existe una correspondencia uno-a-uno entre las formas bilineales simétricas y las formas cuadráticas dadas por la Ecuación (19). De hecho, si K es una forma cuadrática en un espacio vectorial V sobre un campo F cuya característica no sea dos, y si $K(x) = H(x, x)$ para alguna forma bilineal simétrica en V, entonces

$$H(x, y) = \tfrac{1}{2}[K(x + y) - K(x) - K(y)]. \qquad (20)$$

Véase el Ejercicio 15 para más detalles.

Ejemplo 38. Ciertamente, el ejemplo clásico de una forma cuadrática es el del polinomio de segundo grado homogéneo de varias variables. Dadas las variables $t_1, t_2, \ldots, t_n$ que toman valores en un campo F cuya característica no sea dos y los escalares (no necesariamente distintos) $a_{ij}(1 \leq \leq i \leq j \leq n)$, defínase el polinomio

$$f(t_1, t_2, \ldots, t_n) = \sum_{i \leq j} a_{ij} t_i t_j.$$

Sea K: $F^n \to F$ la forma cuadrática definida mediante $K(c_1, c_2, \ldots, c_n) = f(c_1, c_2, \ldots, c_n)$.

Cualquier polinomio de la forma anterior se denomina *polinomio homogéneo de segundo grado en* n *variables.* De hecho, si β es la base estándar para F^n, entonces la forma bilineal simétrica correspondiente a la forma cuadrática anterior es H, donde $\psi_\beta(H) = A$ y

$$A_{ij} = A_{ji} = \begin{cases} a_{ii} & \text{si } i = j \\ \tfrac{1}{2} a_{ij} & \text{si } i \neq j. \end{cases}$$

Para ver esto, simplemente aplíquese la Ecuación (20) para obtener $H(e_i, e_j) = A_{ij}$ a partir de la forma cuadrática K y verifíquese que $f(t_1, \ldots, t_n)$ es calculable a partir de H por medio de la Ecuación (19). En particular, dado el polinomio

$$f(t_1, t_2, t_3) = 2t_1^2 - t_2^2 + 6t_1t_2 - 4t_2t_3$$

con coeficientes reales, sea

$$A = \begin{pmatrix} 2 & 3 & 0 \\ 3 & -1 & -2 \\ 0 & -2 & 0 \end{pmatrix}.$$

Haciendo $H(x, y) = x^t A y$ para toda $x, y \in \mathsf{R}^3$, vemos que

$$f(t_1, t_2, t_3) = (t_1, t_2, t_3) A \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \qquad \text{para} \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \in \mathsf{R}^3.$$

Formas cuadráticas sobre el campo R

Como las matrices simétricas sobre R son "ortogonalmente diagonalizables" (véase el Teorema 7.21), la teoría de las formas bilineales simétricas y de las formas cuadráticas en espacios vectoriales dimensionalmente finitos sobre R es especialmente adecuada. El teorema siguiente y su corolario se encuentran definitivamente entre los resultados más útiles en la teoría de las formas bilineales y cuadráticas.

Teorema 7.35. *Sea* V *un espacio real dimensionalmente finito con producto interior, y sea* H *una forma bilineal simétrica en* V*. Entonces, existe una base ortonormal* β *para* V *tal que* $\psi_\beta(\mathrm{H})$ *es una matriz diagonal.*

DEMOSTRACIÓN. Tómese cualquier base ortonormal $\gamma = \{x_1, x_2, \ldots x_n\}$ para V. Sea $A = \psi_\gamma(H)$. Como A es una matriz simétrica, es autoadjunta con respecto al producto interior en V. Aplicando el Teorema 7.21, podemos encontrar una matriz ortogonal Q tal que $D = Q^tAQ$ sea una matriz diagonal. Para $j = 1, 2, \ldots, n$, defínase

$$y_j = \sum_{i=1}^{n} Q_{ij}x_i.$$

Es una operación sencilla verificar que $\beta = \{y_1, y_2, \ldots, y_n\}$ es una base ortonormal para V. Además, debido a la manera en que β está definida, Q es la matriz de cambio de coordenadas que transforma las coordenadas de β en coordenadas de γ. Por lo tanto, de acuerdo con el Teorema 7.32,

$$\psi_\beta(H) = Q^t\psi\ (H)Q = Q^tAQ = D,$$

la cual es una matriz diagonal. ■

Corolario. *Sea* K *una forma cuadrática en un espacio real con producto interior dimensionalmente finito* V*. Existe una base ortonormal* $\beta = \{\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n\}$ *para* V *y escalares* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *(no necesariamente distintos), tales que si* $\mathrm{x} \in \mathrm{V}$ *y*

$$\mathrm{x} = \sum_{i=1}^{n} s_i x_i, \qquad s_i \in R,$$

entonces

$$K(x) = \sum_{i=1}^{n} \lambda_i s_i^2.$$

De hecho, si H *es la forma bilineal simétrica determinada por* K*, entonces* β *puede escogerse para que sea cualquier base ortonormal para* V *para la cual* $\psi_\beta(\mathrm{H})$ *es una matriz diagonal.*

DEMOSTRACIÓN. Sea H la forma bilineal simétrica para la cual $K(x) = H(x, x)$ para toda $x \in V$. De acuerdo con el Teorema 7.35, existe una base ortonormal $\beta = \{x_1, x_2, \ldots, x_n\}$ para V para la cual

$$\psi_\beta(H) = D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Sea $x \in V$ y supóngase que

$$x = \sum_{i=1}^{n} s_i x_i.$$

Entonces

$$K(x) = H(x, x) = [\phi_\beta(x)]^t D[\phi_\beta(x)] = (s_1, \ldots, s_n) D \begin{pmatrix} s_1 \\ \cdot \\ \cdot \\ \cdot \\ s_n \end{pmatrix} = \sum_{i=1}^{n} \lambda_i s_i^2. \quad \blacksquare$$

Ejemplo 39. Para el polinomio real homogéneo de grado 2

$$f(t_1, t_2) = 5t_1^2 + 2t_2^2 + 4t_1t_2, \tag{21}$$

encontraremos una base ortonormal $\beta = \{x_1, x_2\}$ para R^2 y escalares λ_1 y λ_2 tales que si

$$\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \in R^2 \quad \text{y} \quad \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = s_1x_1 + s_2x_2,$$

entonces $f(t_1, t_2) = \lambda_1 s_1^2 + \lambda_2 s_2^2$. Podemos pensar en s_1 y s_2 como las coordenadas de

$$\begin{pmatrix} t_1 \\ t_2 \end{pmatrix}$$

relativas a β. Así, el polinomio $f(t_1, t_2)$, como expresión que involucra a las coordenadas de un punto con respecto a la base estándar para R^2, se transforma en un nuevo polinomio $g(s_1, s_2) = \lambda_1 s_1^2 + \lambda_2 s_2^2$ interpretado como una expresión que involucra a las coordenadas relativas de un punto a la nueva base β.

Sea H la forma bilineal simétrica correspondiente a la forma cuadrática definida por la Ecuación (21). Si γ es la base estándar para R^2, entonces

$$\psi_\gamma(H) = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}.$$

Ahora encontraremos una matriz ortogonal Q para la cual Q^tAQ sea una matriz diagonal. Como en la Sección 7.7, comenzaremos calculando una base ortonormal de vectores de L_A. El polinomio característico $h(t)$ de A es

$$h(t) = \det\begin{pmatrix} 5-t & 2 \\ 2 & 2-t \end{pmatrix} = (t-6)(t-1).$$

Por lo tanto, $\lambda_1 = 6$ y $\lambda_2 = 1$ son los eigenvalores de A y cada uno de ellos tiene multiplicidad 1. Un cálculo sencillo da los eigenvectores correspondientes de norma uno,

$$x_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \text{y} \quad x_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ -2 \end{pmatrix}.$$

Como x_1 y x_2 son ortogonales, $\beta = \{x_1, x_2\}$ es una base ortonormal para R^2. Haciendo

$$Q = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix},$$

vemos que Q es una matriz ortogonal y

$$Q^tAQ = \begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}.$$

Evidentemente, Q es también la matriz de cambio de coordenadas. Por lo tanto,

$$\psi_\beta(H) = Q^t\psi_\alpha(H)Q = Q^tAQ = \begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}.$$

Entonces, de acuerdo con el corolario del Teorema 7.35, para cualquier $x = s_1x_1 + s_2x_2 \in \mathsf{R}^2$,

$$K(x) = 6s_1^2 + 1s_2^2.$$

Y así $g(s_1, s_2) = 6s_1^2 + 1s_2^2$.

El ejemplo siguiente ilustra cómo puede aplicarse la teoría de las formas cuadráticas al problema de la descripción de las superficies cuadráticas en R^3.

Ejemplo 40. Considérese la superficie $\mathcal{S}$ en R^3 definida por la ecuación

$$2t_1^2 + 6t_1t_2 + 5t_2^2 - 2t_2t_3 + 2t_3^2 + 3t_1 - 2t_2 - t_3 + 14 = 0; \qquad (22)$$

es decir, $\mathcal{S}$ es el conjunto de todos los elementos

$$\begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \in \mathsf{R}^3$$

que satisfacen la Ecuación (22). Si γ es la base estándar para R^3, entonces la Ecuación (22) es una ecuación que involucra a las coordenadas de los puntos en $\mathcal{S}$ relativos a γ. Nos gustaría tomar una nueva base ortonormal β para R^3 tal que la ecuación que describe las coordenadas de cualquier punto de $\mathcal{S}$ relativas a β sea considerablemente más sencilla que la Ecuación (22).

Principiamos con la observación de que los términos de segundo grado en el lado izquierdo de la Ecuación (22) se suman para formar una forma cuadrática en R^3:

$$K\begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} = 2t_1^2 + 6t_1t_2 + 5t_2^2 - 2t_2t_3 + 2t_3^2.$$

En seguida diagonalizamos a K. Si H es la forma bilineal simétrica correspondiente a K y $A = \psi_\gamma(H)$, entonces

$$A = \begin{pmatrix} 2 & 3 & 0 \\ 3 & 5 & -1 \\ 0 & -1 & 2 \end{pmatrix}.$$

Ahora bien, el polinomio característico de A es

$$h(t) = \det\begin{pmatrix} 2-t & 3 & 0 \\ 3 & 5-t & -1 \\ 0 & -1 & 2-t \end{pmatrix} = -1(t-2)(t-7)t,$$

y, por lo tanto, A tiene los eigenvalores $\lambda_1 = 2$, $\lambda_2 = 7$ y $\lambda_3 = 0$. Un cálculo sencillo da los eigenvectores

$$x_1 = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}, \qquad x_2 = \frac{1}{\sqrt{35}}\begin{pmatrix} 3 \\ 5 \\ -1 \end{pmatrix} \quad \text{y} \quad x_3 = \frac{1}{\sqrt{14}}\begin{pmatrix} -3 \\ 2 \\ 1 \end{pmatrix}$$

de norma 1 correspondientes a los eigenvalores respectivos.

Ahora, hágase a $\beta = \{x_1, x_2, x_3\}$ y

$$Q = \begin{pmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{35}} & -\frac{3}{\sqrt{14}} \\ 0 & \frac{5}{\sqrt{35}} & \frac{2}{\sqrt{14}} \\ \frac{3}{\sqrt{10}} & -\frac{1}{\sqrt{35}} & \frac{1}{\sqrt{14}} \end{pmatrix}.$$

Como en el Ejemplo 39, Q es la matriz de cambio de coordenadas que transforma las coordenadas de β en coordenadas de γ y

$$\psi_\beta(H) = Q^t\psi_\gamma(H)Q = Q^tAQ = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Por el corolario del Teorema 7.35, si $x = s_1x_1 + s_2x_2 + s_3x_3$, entonces

$$K(x) = 2s_1^2 + 7s_2^2. \tag{23}$$

Ahora estamos preparados para transformar la Ecuación (22) en una ecuación que contenga coordenadas relativas a β.

Sea

$$x = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \in \mathsf{R}^3.$$

Si $x = s_1x_1 + s_2x_2 + s_3x_3$, tenemos

$$\begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} = s_1 \begin{pmatrix} \frac{1}{\sqrt{10}} \\ 0 \\ \frac{3}{\sqrt{10}} \end{pmatrix} + s_2 \begin{pmatrix} \frac{3}{\sqrt{35}} \\ \frac{5}{\sqrt{35}} \\ \frac{-1}{\sqrt{35}} \end{pmatrix} + s_3 \begin{pmatrix} \frac{-3}{\sqrt{14}} \\ \frac{2}{\sqrt{14}} \\ \frac{1}{\sqrt{14}} \end{pmatrix}.$$

Entonces

$$\begin{aligned} t_1 &= \frac{s_1}{\sqrt{10}} + \frac{3s_2}{\sqrt{35}} - \frac{3s_3}{\sqrt{14}}, \\ t_2 &= \qquad\quad \frac{5s_2}{\sqrt{35}} + \frac{2s_3}{\sqrt{14}}, \\ t_3 &= \frac{3s_1}{\sqrt{10}} - \frac{s_2}{\sqrt{35}} + \frac{s_3}{\sqrt{14}}. \end{aligned}$$

Por lo tanto

$$3t_1 - 2t_2 - t_3 = -\frac{14s_3}{\sqrt{14}} = -\sqrt{14}s_3. \tag{24}$$

Combinando las Ecuaciones (22), (23) y (24), concluimos que si $x \in \mathsf{R}^3$ y $x = s_1x_1 + s_2x_2 + s_3x_3$, entonces $x \in \mathcal{S}$ si y sólo si

$$2s_1^2 + 7s_2^2 - \sqrt{14}s_3 + 14 = 0 \quad \text{o bien} \quad s_3 = \frac{\sqrt{14}}{7}s_1^2 + \frac{\sqrt{14}}{2}s_2^2 + \sqrt{14}. \tag{25}$$

En consecuencia, si establecemos nuevos ejes x', y' y z' en las direcciones x_1, x_2 y x_3, respectivamente, la gráfica de la Ecuación (25) reescrita como

$$z' = \frac{\sqrt{14}}{7}(x')^2 + \frac{\sqrt{14}}{2}(y')^2 + \sqrt{14}$$

coincidirá con la superficie $\mathcal{S}$. Así reconocemos que $\mathcal{S}$ es un paraboloide elíptico. Véase la Figura 7.9.

Concluiremos esta sección con una aplicación de la teoría de formas cuadráticas al cálculo en varias variables —la derivación de la prueba de la segunda derivada para puntos extremos locales de una función de varias variables. Supondremos que el lector posee un conocimiento en el cálculo de funciones de varias variables hasta un nivel del teorema de Taylor. El lector ya está, sin duda, familiarizado con la versión de una variable del teorema de Taylor. Para el enunciado y la demostración de la versión de varias variables, consúltese, por ejemplo, *Advanced Calculus* por Avner Friedman, Holt, Rinehart y Winston, Inc., 1971.

Sea $z = f(t_1, t_2, \ldots, t_n)$ una función de valor real de n variables para la cual todas las derivadas parciales de tercer orden existen y son continuas. Se dice que la función f tiene un *máximo local* en el punto $p \in \mathrm{R}^n$ si existe un número positivo δ para el cual $f(p) \leq f(x)$ siempre que $||x - p|| < \delta$. De la misma manera, se dice que f tiene un *mínimo local* en $p \in \mathrm{R}^n$ si, para algún número $\delta > 0$, $f(p) \leq f(x)$ siempre que $||x - p|| < \delta$. Si f tiene un máximo o un mínimo local en p, decimos que f tiene un *extremo local* en p. Un punto $p \in \mathrm{R}^n$ se llama *punto crítico* de f si $\partial f(p)/\partial t_i = 0$ para $i = 1, 2, \ldots, n$. Es un hecho muy conocido que si f tiene un extremo local en un punto $p \in \mathrm{R}^n$, entonces p es un punto crítico

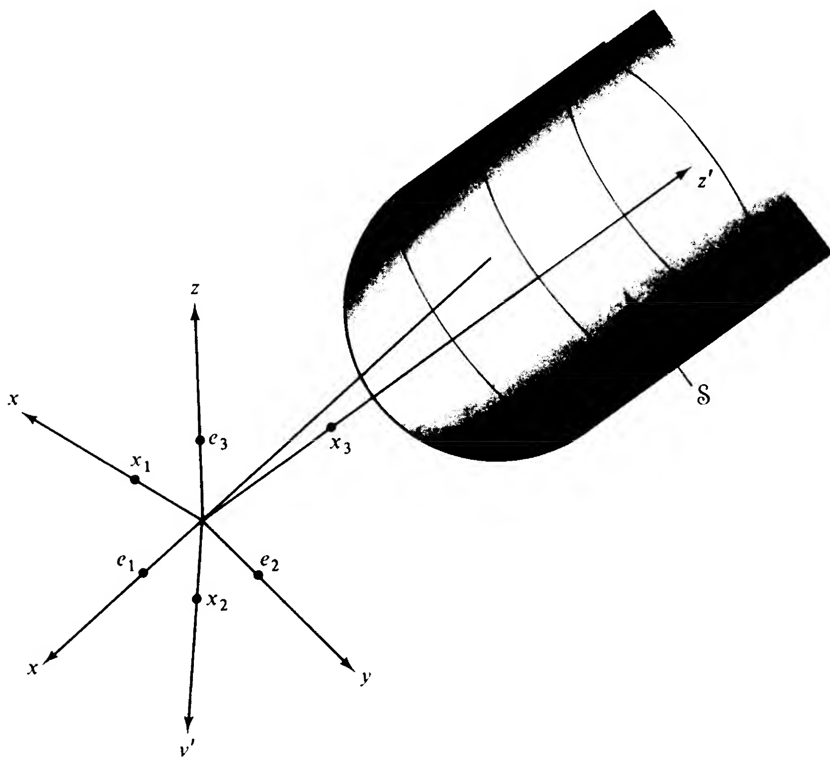


figura 7.9

de f. Ya que, si f tiene un extremo local en p, entonces para cualquier $i = 1, 2, \ldots, n$ podemos definir una función ϕ_i de valor real de una variable mediante $\phi_i(t) = f(p_1, p_2, \ldots, p_{i-1}, t, p_{i+1}, \ldots, p_n)$, donde p_j es la coordenada j de p para cada j. Evidentemente, ϕ_i tiene un extremo local en $t = p_i$. Luego, empleando argumentos ordinarios de cálculo de una variable,

$$\frac{d\phi_i(p_i)}{dt} = \frac{\partial f(p)}{\partial t_i} = 0.$$

Por lo tanto, p es un punto crítico de f. Desafortunadamente, los puntos críticos no son necesariamente extremos locales; pero la prueba de la segunda derivada nos proporciona condiciones adicionales bajo las cuales los puntos críticos son extremos locales.

Teorema 7.36. *(La prueba de la segunda derivada.) Sea* $f(t_1, t_2, \ldots, t_n)$ *una función de valor real de* n *variables para la cual existen todas las derivadas parciales de tercer orden y son continuas. Sea* p *un punto crítico de* f, *y sea* A *la matriz de* $n \times n$ *cuyos elementos estén dados por*

$$A_{ij} = \frac{\partial^2 f(p)}{(\partial t_i)(\partial t_j)}.$$

(Nótese que A *es una matriz simétrica y, por lo tanto, tiene eigenvalores reales.) Entonces*

(a) *Si todos los eigenvalores de* A *son positivos, entonces* f *tiene un mínimo local en* p.

(b) *Si todos los eigenvalores de* A *son negativos, entonces* f *tiene un máximo local en* p.

(c) *Si* A *tiene al menos un eigenvalor positivo y uno negativo, entonces* f *no tiene un extremo local en* p *(esto es,* p *es un punto silla de* f*).*

(d) *Si rango*$(A) < n$ *y* A *no tiene eigenvalores positivos ni negativos, entonces la prueba de la segunda derivada no permite obtener ninguna conclusión.*

DEMOSTRACIÓN. Si $p \neq 0$, podemos definir una función g: $R^n \to R$ mediante

$$g(t_1, t_2, \ldots, t_n) = f(t_1 + p_1, t_2 + p_2, \ldots, t_n + p_n) - f(p_1, p_2, \ldots, p_n).$$

Las siguientes observaciones pueden verificarse fácilmente:

1. La función f tiene un máximo [mínimo] local en p si y sólo si g tiene un máximo [mínimo] local en $0 = (0, 0, \ldots, 0)$.
2. Las derivadas parciales de g en 0 coinciden con las correspondientes derivadas parciales de f en p.

3. O es un punto crítico de g.
4. $A_{ij} = \dfrac{\partial^2 g(O)}{(\partial t_i)(\partial t_j)}$ para toda i y toda j.

En vista de lo anterior podemos suponer sin perder generalidad que $p = O$ y que $f(p) = 0$.

Aplicaremos en seguida el teorema de Taylor a f en O y concluiremos que existe una función de valor real S en R^n tal que

$$\lim_{x\to O} \frac{S(x)}{\|x\|^2} = \lim_{(t_1,\ldots,t_n)\to O} \frac{S(t_1,\ldots,t_n)}{t_1^2+\cdots+t_n^2} = 0, \tag{26}$$

y

$$f(t_1,\ldots,t_n) = f(O) + \sum_{i=1}^{n} \frac{\partial f(O)}{\partial t_i} t_i + \frac{1}{2}\left[\sum_{i,j=1}^{n} \frac{\partial^2 f(O)}{(\partial t_i)(\partial t_j)} t_i t_j\right] + S(t_1,\ldots,t_n). \tag{27}$$

Bajo las hipótesis de que O es un punto crítico y $f(O) = 0$, la Ecuación (27) se reduce a

$$f(t_1,\ldots,t_n) = \frac{1}{2}\left[\sum_{i,j=1}^{n} \frac{\partial^2 f(O)}{(\partial t_i)(\partial t_j)} t_i t_j\right] + S(t_1,\ldots,t_n). \tag{28}$$

Definamos una forma cuadrática K: $\mathsf{R}^n \to R$ mediante

$$K\begin{pmatrix} t_1 \\ \cdot \\ \cdot \\ \cdot \\ t_n \end{pmatrix} = \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 f(O)}{(\partial t_i)(\partial t_j)} t_i t_j. \tag{29}$$

Sea H la forma bilineal simétrica correspondiente a K, y sea γ la base estándar para R^n. Es fácil verificar que $\psi_\gamma(H) = \frac{1}{2}A$. Como A es autoadjunta, el Teorema 7.21 muestra que existe una matriz ortogonal Q tal que

$$Q^t A Q = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

es una matriz diagonal cuyos elementos de la diagonal son los eigenvalores de A. Sea $\beta = \{x_1, x_2, \ldots, x_n\}$ la base ortonormal para R^n cuyo miembro i es Q^i, la columna i de Q. Entonces, Q es la matriz de cambio de

coordenadas que transforma las coordenadas de β en coordenadas de γ y, de acuerdo con el Teorema 7.32,

$$\psi_\beta(H) = Q^t\psi_\gamma(H)Q = \frac{1}{2}Q^tAQ = \begin{pmatrix} \frac{\lambda_1}{2} & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2}{2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{\lambda_n}{2} \end{pmatrix}.$$

Supóngase que A no es la matriz nula. Entonces, A tiene eigenvalores no nulos. Tómese un número positivo ϵ tal que $\epsilon < |\lambda_i|/2$ para todo $\lambda_i \neq 0$. De acuerdo con la Ecuación (26) existe un número positivo δ tal que si $x \in \mathrm{R}^n$ y $||x|| < \delta$, entonces $|S(x)| < \epsilon||x||^2$. Ahora tómese cualquier $x \in \mathrm{R}^n$ para la cual $||x|| < \delta$. Entonces, de acuerdo con las Ecuaciones (28) y (29)

$$|f(x) - K(x)| = |S(x)| < \epsilon||x||^2$$

o bien

$$K(x) - \epsilon||x||^2 < f(x) < K(x) + \epsilon||x||^2. \tag{30}$$

Si

$$x = \sum_{i=1}^{n} s_i x_i,$$

entonces

$$||x||^2 = \sum_{i=1}^{n} s_i^2 \quad \text{y} \quad K(x) = \frac{1}{2}\sum_{i=1}^{n} \lambda_i s_i^2. \tag{31}$$

Luego, de acuerdo con las Ecuaciones (30) y (31)

$$\sum_{i=1}^{n}\left(\frac{1}{2}\lambda_i - \epsilon\right)s_i^2 < f(x) < \sum_{i=1}^{n}\left(\frac{1}{2}\lambda_i + \epsilon\right)s_i^2. \tag{32}$$

Ahora supóngase que todos los eigenvalores de A son *positivos*. Entonces, $\frac{1}{2}\lambda_i - \epsilon > 0$ para toda i y, por lo tanto, por la desigualdad de la izquierda de la Ecuación (32),

$$f(0) = 0 \leq \sum_{i=1}^{n}\left(\frac{1}{2}\lambda_i - \epsilon\right)s_i < f(x).$$

Así, para $||x|| < \delta$, $f(0) \leq f(x)$. Concluimos que f tiene un mínimo local en 0. De la misma manera, mediante un argumento que involucre a la desigualdad derecha de la Ecuación (32), concluimos que si todos los

eigenvalores de A son negativos, entonces f tiene un máximo local en 0. Esto establece los incisos (a) y (b) del teorema.

Ahora supóngase que A tiene un eigenvalor positivo y uno negativo, digamos $\lambda_i > 0$ y $\lambda_j < 0$ para alguna i y alguna j. Entonces $\frac{1}{2}\lambda_i - \epsilon > 0$ y $\frac{1}{2}\lambda_j + \epsilon < 0$. Sea s cualquier número real tal que $|s| < \delta$. Entonces, según la Ecuación (32),

$$f(0) = 0 \leq (\tfrac{1}{2}\lambda_i - \epsilon)s^2 < f(sx_i) \text{ y } f(sx_j) < (\tfrac{1}{2}\lambda_j + \epsilon)s^2 \leq 0 = f(0).$$

Puesto que $\|sx_i\| = \|sx_j\| = |s|$, concluimos que f alcanza valores positivos y negativos arbitrariamente cercanos a *cero*. Por lo tanto, f no tiene ni un máximo local ni un mínimo local en *cero*. Esto demuestra el inciso (c).

Para ilustrar que la prueba de la segunda derivada no permite obtener ninguna conclusión bajo las condiciones establecidas en el inciso (d) del teorema, considérense las funciones

$$f(t_1, t_2) = t_1^2 - t_2^4 \quad \text{y} \quad f(t_1, t_2) = t_1^2 + t_2^4$$

en $p = 0$. En ambos casos

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix},$$

pero en el primer caso f no tiene un extremo local en 0, mientras que en el último caso f tiene un mínimo local en 0. ■

EJERCICIOS

1. Decir si las siguientes afirmaciones son verdaderas o falsas.
   (a) Toda forma cuadrática es una forma bilineal.
   (b) Si dos matrices son congruentes, tienen los mismos eigenvalores.
   (c) Las formas bilineales simétricas tienen representaciones matriciales simétricas.
   (d) Cualquier matriz simétrica es congruente con una matriz diagonal.
   (e) La suma de dos formas bilineales simétricas es una forma bilineal simétrica.
   (f) Dos matrices simétricas con el mismo polinomio característico son representaciones matriciales de la misma forma bilineal.
   (g) Existe una forma bilineal H tal que $H(x, y) \neq 0$ para toda x y toda y.
   (h) Si V es un espacio vectorial de dimensión n, entonces $\dim(\mathcal{B}(V)) = 2n$.
   (i) Sea H una forma bilineal en un espacio vectorial dimensionalmente finito V. Para cualquier $x \in V$ existe una $y \in V$ tal que $y \neq 0$ pero $H(x, y) = 0$.
   (j) Si H es una forma bilineal cualquiera en un espacio real V dimensionalmente finito, con producto interior entonces existe una base β para V tal que $\psi_\beta(H)$ es una matriz diagonal.

2. Demostrar las propiedades 1, 2, 3 y 4 dadas en la p. 469.

3. (a) Verificar que la suma de dos formas bilineales es una forma bilineal.
   (b) Verificar que el producto de un escalar por una forma bilineal es una forma bilineal.
   (c) Demostrar el Teorema 7.30.

4. Determinar en cuál de los incisos siguientes se tienen formas bilineales.
   (a) Sea V = C[0, 1] el espacio de las funciones continuas de valor real en el intervalo cerrado [0, 1]. Para f, $g \in$ V, defínase

$$H(f, g) = \int_0^1 f(t)g(t)dt.$$

   (b) Sea V un espacio vectorial sobre un campo F, y sea $J \in \mathcal{B}(V)$ no trivial. Defínase a H: $V \times V \to F$ mediante

$$H(x, y) = [J(x, y)]^2 \qquad \text{para toda } x, y \in V.$$

   (c) Defínase a H: $R \times R \to R$ mediante $H(t_1, t_2) = t_1 + 2t_2$.
   (d) Considérense los miembros de R^2 como vectores columnas. Definir a H: $R^2 \to R$ para x, $y \in R^2$ mediante $H(x, y) = \det(x, y)$, donde $\det(x, y)$ representa el determinante de la matriz de 2×2 con x como su primera columna y y como su segunda columna.
   (e) Sea V un espacio real con producto interior. Defínase a H: $V \times V \to R$ mediante $H(x, y) = (x, y)$ para x, $y \in V$.
   (f) Sea V un espacio complejo con producto interior. Defínase a H: $V \times V \to C$ mediante $H(x, y) = (x, y)$ para x, $y \in V$.

5. Verificar que cada uno de los mapeos dados es una forma bilineal. Luego calcular la representación matricial de H con respecto a la base dada.
   (a) H: $R^3 \times R^3 \to R$, donde

$$H\left(\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}\right) = a_1b_1 - 2a_1b_2 + a_2b_1 - a_3b_3$$

   con

$$\beta = \left\{\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right\}.$$

   (b) Sea V $= M_{2\times 2}(R)$ con la base

$$\beta = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

   Defínase a H: $V \times V \to R$ mediante $H(A, B) = \operatorname{tr}(A) \cdot \operatorname{tr}(B)$.

(c) Sea $\beta = \{\cos t, \text{ sen } t, \cos 2t, \text{ sen } 2t\}$ y $V = L(\beta)$. En el espacio de todas las funciones continuas en R, V es un espacio tetradimensional con una base β. Defínase a H: $V \times V \rightarrow R$ mediante $H(f, g) = f'(0) \cdot g''(0)$.

6. Sean V y W espacios vectoriales sobre el mismo campo, y sea T: $V \rightarrow W$ una transformación lineal. Para cualquier $H \in \mathcal{B}(W)$, defínase a $T(H)$: $V \times V \longrightarrow F$ mediante $\hat{T}(H)(x, y) = H(T(x), T(y))$ para toda $x, y \in V$. Demostrar que

(a) Para $H \in \mathcal{B}(W)$, $\hat{T}(H) \in \mathcal{B}(V)$.
(b) T: $\mathcal{B}(W) \longrightarrow \mathcal{B}(V)$ es una transformación lineal.
(c) Si T es un isomorfismo, $\hat{T}$ también lo es.

7. En la demostración del Teorema 7.31

(a) Demostrar que para cualquier base β, ψ_β es lineal.
(b) Sea V un espacio vectorial n-dimensional sobre un campo F con una base β, y sea ϕ_β: $V \rightarrow F^n$ la representación ordinaria de V con respecto a β. Sea $A \in M_{n \times n}(F)$. Defínase a H: $V \times V \rightarrow F$ mediante $H(x, y) = [\phi_\beta(x)]^t A[\phi_\beta(y)]$. Demostrar que $H \in \mathcal{B}(V)$. ¿Se podría establecer esto como corolario al Ejercicio 6?

8. (a) Demostrar el Corolario 1 del Teorema 7.31.
(b) Para un espacio vectorial dimensionalmente finito V, describa un método para encontrar una base para $\mathcal{B}(V)$.

9. Demostrar el Corolario 2 del Teorema 7.31.

10. Demostrar el Corolario 3 del Teorema 7.31.

11. Demostrar que la relación de congruencia es una relación de equivalencia.

12. La siguiente descripción nos proporciona una demostración alternativa para el Teorema 7.32.

(a) Si β y γ son bases para un espacio vectorial dimensionalmente finito V, y si Q es la matriz de cambio de coordenadas, demostrar que $\phi_\beta = L_Q\phi_\gamma$, donde ϕ_β y ϕ_γ son las representaciones ordinarias de V con respecto a β y γ, respectivamente.
(b) Aplicar el Corolario 2 del Teorema 7.31 al inciso (a) para obtener una demostración alternativa del Teorema 7.32.

13. Sea V un espacio vectorial dimensionalmente finito y $H \in \mathcal{B}(V)$. Demostrar que, para bases cualesquiera β y γ de V, $\text{rango}(\psi_\beta(H)) = \text{rango}(\psi_\gamma(H))$.

14. Demostrar las proposiciones siguientes.

(a) Cualquier matriz cuadrada diagonal es simétrica.
(b) Cualquier matriz congruente con una matriz diagonal es simétrica.
(c) Demostrar el corolario del Teorema 7.34.

15. Sea V un espacio vectorial sobre un campo F cuya característica no sea dos, y sea H una forma bilineal simétrica en V. Demostrar que si $K(x) = H(x, x)$ es la forma cuadrática asociada con H, entonces

$$H(x, y) = \tfrac{1}{2}[K(x + y) - K(x) - K(y)]$$

para toda $x, y \in V$.

16. Para las siguientes formas cuadráticas K sobre un espacio real con producto interior V, encontrar una forma bilineal simétrica H tal que $K(x) = H(x, x)$ para toda $x \in V$. Después, encontrar una base β ortonormal para V tal que $\psi_\beta(H)$ sea una matriz diagonal.

(a) K: $R^2 \to R$ definida mediante

$$K\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = -2t_1^2 + 4t_1t_2 + t_2^2$$

(b) K: $R^2 \to R$ definida mediante

$$K\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = 7t_1^2 - 8t_1t_2 + t_2^2$$

(c) K: $R^3 \to R$ definida mediante

$$K\begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} = 3t_1^2 + 3t_2^2 + 3t_3^2 - 2t_1t_3$$

17. Sea $\mathcal{S}$ el conjunto de todos los

$$\begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix} \in R^3$$

tales que

$$3t_1^2 + 3t_2^2 + 3t_3^2 - 2t_1t_3 + 2\sqrt{2}(t_1 + t_3) + 1 = 0.$$

Encontrar una base ortonormal β para R^3 tal que se simplifique la ecuación que relaciona las coordenadas de los puntos de $\mathcal{S}$ relativas a β. Describir geométricamente a $\mathcal{S}$.

18. Demostrar la siguiente expresión más refinada del inciso (d) del Teorema 7.36.

(a) Si $0 < \text{rango}(A) < n$ y A tiene eigenvalores no negativos, entonces f no tiene un mínimo local en p.
(b) Si $0 < \text{rango}(A) < n$ y A tiene eigenvalores no positivos, entonces f no tiene un máximo local en p.

19. Demostrar la siguiente variación de la prueba de la segunda derivada para el caso en que $n = 2$. Defínase a

$$D = \left[\frac{\partial^2 f(p)}{(\partial t_1)^2}\right]\left[\frac{\partial^2 f(p)}{(\partial t_2)^2}\right] - \left[\frac{\partial^2 f(p)}{(\partial t_1)(\partial t_2)}\right]^2.$$

(a) Si $D > 0$ y $\partial^2 f(p)/(\partial t_1)^2 > 0$, entonces f tiene un mínimo local en p.
(b) Si $D > 0$ y $\partial^2 f(p)/(\partial t_2)^2 < 0$, entonces f tiene un máximo local en p.
(c) Si $D < 0$, entonces f no tiene un extremo local en p.
(d) Si $D = 0$, entonces no se puede obtener conclusión alguna.

Sugerencia: Obsérvese que $D = \det(A) = \lambda_1\lambda_2$, donde λ_1 y λ_2 son los eigenvalores de A, y A es como en el Teorema 7.36.

20. Sea A una matriz de $n \times n$ sobre un campo F, y sea E una matriz elemental de $n \times n$ sobre F. En la Sección 3.1 se demostró que AE puede obtenerse a partir de A mediante una operación elemental con columnas. Demostrar que E^tA puede obtenerse a partir de A mediante la misma operación elemental, pero realizada sobre los renglones y no sobre las columnas de A. *Sugerencia:* Nótese que $E^tA = (A^tE)^t$.

21. Para cada una de las siguientes matrices A con elementos del campo de los números racionales, encontrar una matriz diagonal D y una matriz invertible Q tal que $Q^tAQ = D$.

(a) $A = \begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix}$

(b) $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

Sugerencia: Utilizar una operación elemental distinta de la de intercambiar columnas.

(c) $A = \begin{pmatrix} 3 & 1 & 2 \\ 1 & 4 & 0 \\ 2 & 0 & -1 \end{pmatrix}$

INDICE DE DEFINICIONES PARA EL CAPITULO 7

Apéndices

APENDICE A CONJUNTOS

Un *conjunto* es una colección de objetos, llamados *elementos* del conjunto. Si x es un elemento del conjunto A, escribimos $x \in A$; si x no es elemento de A, escribimos $x \notin A$. Por ejemplo, si Z es el conjunto de los enteros, entonces $3 \in Z$ y $\frac{1}{2} \notin Z$.

Se dice que dos conjuntos A y B son *iguales,* lo que se escribe como $A = B$, si contienen exactamente los mismos elementos. Los conjuntos se pueden describir de dos maneras:

1. Enlistando todos los elementos del conjunto entre llaves $\{\quad\}$.
2. Describiendo los elementos del conjunto en términos de alguna propiedad característica.

Por ejemplo, el conjunto que consta de los elementos 1, 2, 3 y 4 se puede escribir como $\{1, 2, 3, 4\}$ o como

$$\{x\colon\ x \text{ es un entero positivo menor que } 5\}.$$

Nótese que el orden en el que se enumeran los elementos es intrascendente; por lo tanto

$$\{1, 2, 3, 4\} = \{3, 1, 2, 4\} = \{1, 3, 1, 4, 2\}.$$

Ejemplo 1. Sea A el conjunto de números reales comprendidos entre 1 y 2. Entonces, A puede escribirse como

$$A = \{x\colon\ x \text{ es un número real y } 1 < x < 2\}$$

o bien, si R es el conjunto de los números reales, como

$$A = \{x \in R\colon\ 1 < x < 2\}.$$

Se dice que un conjunto B es *subconjunto* de un conjunto A, lo que se escribe $B \subseteq A$ o $A \supseteq B$, si todo elemento de B es un elemento de A. Por ejemplo, $\{1, 2, 6\} \subseteq \{2, 8, 7, 6, 1\}$. Obsérvese que $A = B$ si y sólo

si $A \subseteq B$ y $B \subseteq A$, un hecho que se utiliza frecuentemente para demostrar que dos conjuntos son iguales.

El *conjunto vacío*, denotado por $\varnothing$, es el conjunto que no contiene ningún elemento. El conjunto vacío es un subconjunto de todo conjunto.

Los conjuntos pueden combinarse para formar otros conjuntos de dos maneras básicas. La *unión* de dos conjuntos A y B, que se escribe $A \cup B$, es el conjunto de los elementos que están en A, o en B, o en ambos; esto es,

$$A \cup B = \{x\colon \ x \in A \text{ o } x \in B\}.$$

La *intersección* de dos conjuntos A y B, que se escribe $A \cap B$, es el conjunto de los elementos que están en A y en B; esto es,

$$A \cap B = \{x\colon \ x \in A \text{ y } x \in B\}.$$

Dos conjuntos se llaman *disjuntos* si su intersección es el conjunto vacío.

Ejemplo 2. Sea $A = \{1, 3, 5\}$ y $B = \{1, 5, 7, 8\}$. Entonces

$$A \cup B = \{1, 3, 5, 7, 8\} \quad \text{y} \quad A \cap B = \{1, 5\}.$$

De manera semejante, si $X = \{1, 2, 8\}$ y $Y = \{3, 4, 5\}$, entonces

$$X \cup Y = \{1, 2, 3, 4, 5, 8\} \quad \text{y} \quad X \cap Y = \varnothing.$$

Por lo tanto, X y Y son conjuntos disjuntos.

La unión y la intersección de más de dos conjuntos puede definirse de una manera análoga. Específicamente, si $A_1, A_2, \ldots, A_n$ son conjuntos, entonces la unión y la intersección de estos conjuntos se define como

$$\bigcup_{i=1}^{n} A_i = \{x\colon \ x \in A_i \text{ para alguna } i = 1, 2, \ldots, n\}$$

y

$$\bigcap_{i=1}^{n} A_i = \{x\colon \ x \in A_i \text{ para toda } i = 1, 2, \ldots, n\}.$$

De una manera semejante, si Λ es un conjunto de índices y $\{A_\alpha\colon \ \alpha \in \Lambda\}$ es una colección de conjuntos, la unión y la intersección de estos conjuntos se definen como

$$\bigcup_{\alpha \in \Lambda} A_\alpha = \{x\colon \ x \in A_\alpha \text{ para alguna } \alpha \in \Lambda\}$$

y

$$\bigcap_{\alpha \in \Lambda} A_\alpha = \{x\colon \ x \in A_\alpha \text{ para toda } \alpha \in \Lambda\}.$$

Ejemplo 3. Sea $\Lambda = \{\alpha \in R: \alpha > 1\}$ y sea

$$A_\alpha = \left\{x \in R: \frac{-1}{\alpha} \leq x \leq 1 + \alpha\right\}$$

para toda $\alpha \in \Lambda$, donde R es el conjunto de números reales. Entonces

$$\bigcup_{\alpha \in \Lambda} A_\alpha = \{x \in R: x > -1\} \text{ y } \bigcap_{\alpha \in \Lambda} A_\alpha = \{x \in R: 0 \leq x \leq 2\}.$$

Por relación en un conjunto A queremos decir que se trata de una regla para determinar, para elementos arbitrarios x e y de A, si x se encuentra o no relacionada con y. Más precisamente, una *relación* en A es un conjunto S de pares ordenados de elementos de A tales que $(x, y) \in S$ si y sólo si x tiene algún parentesco dado con y. Por ejemplo, en el conjunto de números reales, "es igual a", "es menor que" y "es mayor que o igual a" son relaciones comunes. Una relación S en un conjunto A se llama *relación de equivalencia* en A si se cumplen las tres condiciones siguientes:

1. Para toda $x \in A$, $(x, x) \in S$ *(reflexividad).*
2. Si $(x, y) \in S$, entonces $(y, x) \in S$ *(simetría).*
3. Si $(x, y) \in S$ y $(y, z) \in S$, entonces $(x, z) \in S$ *(transitividad).*

Si S es una relación de equivalencia en un conjunto A, escribiremos comúnmente $x \sim y$ en lugar de $(x, y) \in S$. Por ejemplo, si definimos $x \sim y$ para indicar que $x - y$ es divisible entre un entero fijo n, entonces $\sim$ es una relación de equivalencia en el conjunto de los enteros.

APENDICE B FUNCIONES

Si A y B son conjuntos, entonces una *función* f de A en B, que puede escribirse como $f: A \to B$, es una regla que asocia a cada elemento de x en A un elemento único llamado $f(x)$ en B. El elemento $f(x)$ se llama *imagen de* x *(bajo* f) y x se llama *preimagen de* f(x) *(bajo* f). Si $f: A \to B$, entonces A se llama *dominio* de f, y el conjunto $\{f(x): x \in A\}$ de todas las imágenes de los elementos de A se llama *rango* de f. Nótese que el rango de f es un subconjunto de B. Si $S \subseteq A$, denotaremos por $f(S)$ al conjunto $\{f(x): x \in S\}$ de todas las imágenes de los elementos de S. De la misma forma, si $T \subseteq B$, denotaremos por $f^{-1}(T)$ al conjunto $\{x \in A: f(x) \in T\}$ de todas las preimágenes de los elementos de T. Finalmente, dos funciones $f: A \to B$ y $g: A \to B$ son *iguales* si $f(x) = g(x)$ para toda $x \in A$.

Ejemplo 1. Supóngase que $A = [-10, 10]$ y $B = R$, el conjunto de los números reales. Sea $f: A \to B$ la función que asigna a cada elemento x en A el elemento $x^2 + 1$ en B; esto es, f está definida mediante $f(x) =$

$= x^2 + 1$. Entonces, A es el dominio de f y $[1, 101]$ es el rango de f. Como $f(2) = 5$, la imagen de 2 es 5 y 2 es una preimagen de 5. Nótese que -2 es otra preimagen de 5. Más aún, si $S = [1, 2]$ y $T = [82, 101]$, entonces $f(S) = [2, 5]$ y $f^{-1}(T) = [-10, -9] \cup [9, 10]$.

Tal como lo muestra el ejemplo anterior, la preimagen de un elemento del rango no necesariamente es única. Las funciones tales que cada elemento del rango tiene una preimagen única se llama *uno-a-uno*; es decir, $f\colon\ A \to B$ es uno-a-uno si $f(x) = f(y)$ implica que $x = y$ o, de un modo equivalente, si $x \neq y$ implica que $f(x) \neq f(y)$.

Si $f\colon\ A \to B$ es una función de rango B, o sea, si $f(A) = B$, entonces f se llama *sobreyectiva.*

Supóngase que $f\colon\ A \to B$ es una función y $S \subseteq A$. Entonces puede formarse una función $f_S\colon\ S \to B$, que se denomina *restricción de* f *a* S. definiendo $f_S(x) = f(x)$ para cada $x \in S$.

El ejemplo siguiente ilustra estos conceptos.

Ejemplo 2. Sea $f\colon\ [-1, 1] \to [0, 1]$ definida mediante $f(x) = x^2$. Esta función es sobreyectiva pero no uno-a-uno ya que $f(-1) = f(1) = 1$. Nótese que si $S = [0, 1]$, entonces f_S es sobreyectiva y uno-a-uno. Por último, si $T = [\frac{1}{2}, 1]$, entonces f_T es uno-a-uno, pero no sobreyectiva.

Sean A, B y C conjuntos y $f\colon\ A \to B$ y $g\colon\ B \to C$ funciones. Aplicando f seguida de g obtenemos una función $g \circ f\colon\ A \to C$, llamada la *función compuesta* de g y f. Entonces, $(g \circ f)(x) = g(f(x))$ para toda $x \in A$. Por ejemplo, sean $A = B = C = R$ (el conjunto de los números reales), $f(x) = \operatorname{sen} x$ y $g(x) = x^2 + 3$. Entonces, $(g \circ f)(x) = g(f(x)) = \operatorname{sen}^2 x + 3$, mientras que $(f \circ g)(x) = f(g(x)) \operatorname{sen} = (x^2 + 3)$. Por lo tanto, $g \circ f \neq f \circ g$. Sin embargo, la composición funcional es asociativa; esto es, si $h\colon\ C \to D$, entonces $h \circ (g \circ f) = (h \circ g) \circ f$.

Se dice que una función $f\colon\ A \to B$ es *invertible* si existe una función $g\colon\ B \to A$ tal que $(f \circ g)(y) = y$ para toda $y \in B$ y $(g \circ f)(x) = x$ para toda $x \in A$. Si tal función g existe, entonces es única y se llama la *inversa* de f. Escribiremos la inversa de f (cuando exista) como f^{-1}. Puede demostrarse que f es invertible si y sólo si f es sobreyectiva y uno-a-uno.

Ejemplo 3. La función $f\colon\ R \to R$ definida mediante $f(x) = 3x + 1$ es uno-a-uno y sobreyectiva; por lo tanto, es invertible. La inversa de f es la función $f^{-1}\colon\ R \to R$ definida mediante $f^{-1}(x) = (x - 1)/3$.

Los siguientes hechos acerca de las funciones invertibles pueden demostrarse fácilmente:

1. Si $f\colon\ A \to B$ es invertible, entonces f^{-1} es invertible y $(f^{-1})^{-1} = f$.
2. Si $f\colon\ A \to B$ y $g\colon\ B \to C$ son invertibles, entonces $g \circ f$ es invertible y $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

APENDICE C CAMPOS

El conjunto de números reales es un ejemplo de una estructura algebraica llamada "campo". Básicamente, un campo es un conjunto en el cual se pueden definir cuatro operaciones (llamadas adición, multiplicación, substracción y división) tales que, con excepción de la división entre cero, la suma, el producto, la diferencia y el cociente de cualquier par de elementos del conjunto, es un elemento del conjunto. Más detalladamente, un campo se define de la siguiente manera.

Definiciones. *Un campo* F *es un conjunto en el cual se definen dos operaciones* $+$ *y* $\cdot$ *(llamadas, respectivamente,* adición *y* multiplicación*) de modo que para cualquier par de elementos* a, b *en* F *existen elementos únicos* $a + b$ *y* $a \cdot b$ *en* F *tales que se cumplen las siguientes condiciones para todos los elementos* a, b, c *en* F.

(F 1) $a + b = b + a$ *y* $a \cdot b = b \cdot a$
(conmutatividad de la adición y la multiplicación).

(F 2) $(a + b) + c = a + (b + c)$ *y* $(a \cdot b) \cdot c = a \cdot (b \cdot c)$
(asociatividad de la adición y la multiplicación).

(F 3) *Existen elementos distintos* 0 *y* 1 *en* F *tales que*

$$0 + a = a \quad y \quad 1 \cdot a = a$$

(existencia de elementos identidad para la adición y la multiplicación).

(F 4) *Para cada elemento* a *en* F *y cada elemento no nulo* b *en* F *existen elementos* c *y* d *en* F *tales que*

$$a + c = 0 \quad y \quad b \cdot d = 1$$

(existencia de inversos para la adición y la multiplicación).

(F 5) $a \cdot (b + c) = a \cdot b + a \cdot c$
(distributividad de la multiplicación sobre la adición).

Los elementos $a + b$ *y* $a \cdot b$ *se llaman, respectivamente,* suma *y* producto *de* a *y* b. *Los elementos* 0 *(léase* "cero"*)* *y* 1 *(léase* "uno"*) mencionados en* (F 3) *se llaman* elementos de identidad *para la adición y la multiplicación, respectivamente, y los elementos* c *y* d *citados en* (F 4) *se denominan, respectivamente,* inverso aditivo *para* a *e* inverso multiplicativo *para* b.

Ejemplo 1. El conjunto de números reales con las definiciones ordinarias de adición y multiplicación es un campo que se denotará por *R*.

Ejemplo 2. El conjunto de los números racionales con las definiciones ordinarias de adición y multiplicación es un campo.

Ejemplo 3. El conjunto de todos los números reales de la forma $a + b\sqrt{2}$, donde a y b son números racionales, con la adición y la multiplicación como en R, es un campo.

Ejemplo 4. El campo Z_2 consta de dos elementos 0 y 1 con las operaciones de adición y multiplicación definidas por las ecuaciones

$$0 + 0 = 0, \quad 0 + 1 = 1 + 0 = 1, \quad 1 + 1 = 0,$$
$$0 \cdot 0 = 0, \quad 0 \cdot 1 = 1 \cdot 0 = 0, \quad \text{y} \quad 1 \cdot 1 = 1.$$

Ejemplo 5. Ni el conjunto de los enteros positivos ni el de los enteros con las definiciones ordinarias de adición y multiplicación es un campo, puesto que en ambos (F 4) no se satisface.

Los elementos de un campo cuya existencia queda garantizada por (F 3) y (F 4) son únicos; esto es consecuencia del teorema siguiente.

Teorema C.1. *(Leyes de cancelación.) Sean* a, b *y* c *elementos cualesquiera de un campo* F.

(a) *Si* $a + b = c + b$, *entonces* $a = c$.
(b) *Si* $a \cdot b = c \cdot b$ *y* $b \neq 0$, *entonces* $a = c$.

DEMOSTRACIÓN. Las demostraciones de (a) y (b) son semejantes, por lo que únicamente se demostrará (b).

Si $b \neq 0$, entonces (F 4) garantiza la existencia de un elemento d en F tal que $b \cdot d = 1$. Multiplíquense ambos lados de la igualdad $a \cdot b = c \cdot b$ por d para obtener $(a \cdot b) \cdot d = (c \cdot b) \cdot d$. Considérese el lado izquierdo de la igualdad: en virtud de (F 2) y (F 3) tenemos

$$(a \cdot b) \cdot d = a \cdot (b \cdot d) = a \cdot 1 = a.$$

De igual manera, el miembro derecho de la igualdad se reduce a c. Entonces

$$a = (a \cdot b) \cdot d = (c \cdot b) \cdot d = c. \quad \blacksquare$$

Corolario. *Los elementos* 0 *y* 1 *mencionados en* (F 3) *y los elementos* c *y* d *mencionados en* (F 4) *son únicos.*

DEMOSTRACIÓN. Supóngase que $0' \in F$ satisface que $0' + a = a$ para cada $a \in F$. Como $0 + a = a$ para cada $a \in F$, tenemos que $0' + a = 0 + a$ para cada $a \in F$. Por tanto, por el Teorema C.1, $0' = 0$.

La demostración de las partes restantes es similar. $\blacksquare$

Así, cada elemento b en un campo tiene un inverso aditivo único y, si $b \neq 0$, también un inverso multiplicativo único. (Se demostrará en el corolario del Teorema C.2 que 0 no tiene un inverso multiplicativo.)

El inverso aditivo y el inverso multiplicativo de b se escriben $-b$ y b^{-1}, respectivamente. Nótese que $-(-b) = b$ y $(b^{-1})^{-1} = b$.

La substracción y la división se pueden definir en términos de la adición y la multiplicación utilizando los inversos aditivo y multiplicativo. Específicamente, la substracción de b se define como la adición de $-b$ y la división entre $b \neq 0$ se define como la multiplicación por b^{-1}; esto es,

$$a - b = a + (-b) \quad \text{y} \quad a/b = a \cdot b^{-1}.$$

La división entre cero es indefinida, pero, con esta excepción, la suma, el producto, la diferencia y el cociente están definidos para cualquier par de elementos de un campo.

Muchas de las propiedades ordinarias de la multiplicación de números reales son ciertas en cualquier campo, como lo demuestra el teorema siguiente.

Teorema C.2. *Sean* a *y* b *elementos cualesquiera de un campo. Entonces es cierto cada uno de los incisos siguientes.*

(a) $a \cdot 0 = 0$
(b) $(-a) \cdot b = a \cdot (-b) = -(a \cdot b)$
(c) $(-a) \cdot (-b) = a \cdot b$

DEMOSTRACIÓN.

(a) Como $0 + 0 = 0$, (F 5) muestra que

$$a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0.$$

Luego, $0 + a \cdot 0 = a \cdot 0 + a \cdot 0$, y eliminando $a \cdot 0$ por el Teorema C.1 se tiene $0 = a \cdot 0$.

(b) Por definición $-(a \cdot b)$ es el único elemento de F tal que $a \cdot b + [-(a \cdot b)] = 0$. Entonces, con objeto de demostrar que $(-a) \cdot b = -(a \cdot b)$ es suficiente con mostrar que $a \cdot b + (-a) \cdot b = 0$. Pero $-a$ es el elemento de F tal que $a + (-a) = 0$, y entonces

$$a \cdot b + (-a) \cdot b = [a + (-a)] \cdot b = 0 \cdot b = b \cdot 0 = 0$$

por (F 5) y el inciso (a). Así, $(-a) \cdot b = -(a \cdot b)$. La demostración de que $a \cdot (-b) = -(a \cdot b)$ es similar.

(c) Aplicando dos veces el inciso (b), tenemos que

$$(-a) \cdot (-b) = -[a \cdot (-b)] = [-(a \cdot b)] = a \cdot b. \quad ■$$

Corolario. *La identidad aditiva de un campo no tiene inverso multiplicativo.*

En un campo cualquiera F, puede suceder que una suma $1 + 1 + \ldots + 1$ (p sumandos) sea igual a cero para algún entero positivo p. Por ejemplo, en el campo Z_2 (definido en el Ejemplo 4), $1 + 1 = 0$.

En este caso el entero positivo más pequeño posible p para el cual una suma de p 1's es igual a cero, se llama *característica de* F; si no existe tal entero positivo, se dice que F tiene *característica cero.* Así pues, Z_2 tiene característica dos, y R tiene característica cero. Obsérvese que si F es un campo de característica $p \neq 0$, entonces $x + x + \ldots + x$ (p sumandos) es igual a 0 para toda $x \in F$. En un campo de característica finita (especialmente de característica dos) surgen muchos problemas no usuales. Por esta razón, algunos de los resultados sobre espacios vectoriales enunciados en este libro requieren que el campo sobre el que se defina el espacio vectorial sea de característica cero (o, al menos, de alguna característica diferente de dos).

Finalmente, nótese que en otras secciones de este libro, el producto de dos elementos a y b de un campo se expresa ab en vez de $a \cdot b$.

APENDICE D NUMEROS COMPLEJOS

Para los propósitos del álgebra, el campo de los números reales no es suficiente, puesto que existen polinomios de grado no nulo con coeficientes reales que no tienen ceros en el campo de los números reales (por ejemplo, $x^2 + 1$). Es pues a menudo deseable tener un campo en el que cualquier polinomio de grado no nulo con coeficientes de dicho campo tenga un cero en éste. Por esta razón "agrandaremos" el campo de los números reales para obtener tal campo.

Definiciones. *Un* número complejo *es una expresión de la forma* $z = a + bi$, *donde* a *y* b *son números reales llamados, respectivamente,* parte real *y* parte imaginaria *de* z.

La suma *y el* producto *de dos números complejos* $z = a + bi$ *y* $w = c + di$ *(donde* a, b, c *y* d *son números reales) se definen de la siguiente manera:*

$$z + w = (a + bi) + (c + di) = (a + c) + (b + d)i$$

y

$$zw = (a + bi)(c + di) = (ac - bd) + (bc + ad)i.$$

Ejemplo 1. La suma y el producto de $z = 3 - 5i$ y $w = 9 + 7i$ son

$$z + w = (3 - 5i) + (9 + 7i) = (3 + 9) + [(-5) + 7]i = 12 + 2i$$

y

$$zw = (3 - 5i)(9 + 7i) = [3 \cdot 9 - (-5) \cdot 7] + [(-5) \cdot 9 + 3 \cdot 7]i$$
$$= 62 - 24i.$$

Cualquier número real c puede ser considerado como un número complejo asociando a c con el complejo $c + 0i$. Obsérvese que esta correspondencia conserva las sumas y los productos; es decir,

$$(c + 0i) + (d + 0i) = (c + d) + 0i, \quad \text{y} \quad (c + 0i)(d + 0i) = cd + 0i.$$

Cualquier número complejo de la forma $bi = 0 + bi$, donde b es un número real no nulo, se llama *imaginario*. El producto de dos números imaginarios es real puesto que

$$\begin{aligned}(bi)(di) &= (0 + bi)(0 + di) = (0 - bd) + (b \cdot 0 + 0 \cdot d)i \\ &= -bd.\end{aligned}$$

En particular, para $i = 0 + 1i$, tenemos que $i \cdot i = -1$.

La observación de que $i^2 = i \cdot i = -1$ proporciona una manera fácil de recordar la definición de multiplicación de números complejos: sencillamente multiplíquense dos números complejos como se multiplicarían dos expresiones algebraicas y sustitúyase i^2 por -1. El Ejemplo 2 ilustra esta técnica.

Ejemplo 2. El producto de $-5 + 2i$ y $1 - 3i$ es

$$\begin{aligned}(-5 + 2i)(1 - 3i) &= -5(1 - 3i) + 2i(1 - 3i) \\ &= -5 + 15i + 2i - 6i^2 \\ &= -5 + 15i + 2i - 6(-1) \\ &= 1 + 17i.\end{aligned}$$

El número real 0, considerado como un número complejo, es un elemento identidad aditivo para el conjunto de números complejos, puesto que

$$\begin{aligned}(a + bi) + 0 &= (a + bi) + (0 + 0i) = (a + 0) + (b + 0)i \\ &= a + bi.\end{aligned}$$

De una manera análoga, el número real 1, considerado como un número complejo, es un elemento identidad multiplicativo para el conjunto de los números complejos, pues

$$\begin{aligned}(a + bi) \cdot 1 &= (a + bi)(1 + 0i) = (a \cdot 1 - b \cdot 0) + (b \cdot 1 - a \cdot 0)i \\ &= a + bi.\end{aligned}$$

Es evidente que todo número complejo $a + bi$ tiene un inverso aditivo, que es $(-a) + (-b)i$. Pero también todo número complejo, con excepción del 0, tiene un inverso multiplicativo. De hecho,

$$(a + bi)^{-1} = \left(\frac{a}{a^2 + b^2}\right) - \left(\frac{b}{a^2 + b^2}\right)i.$$

En vista de los enunciados anteriores no debe sorprendernos el resultado siguiente.

Teorema D.1. *El conjunto de números complejos con las operaciones de adición y multiplicación definidas anteriormente es un campo.*

El campo de los números complejos será representado por C.

Definición. *El* (complejo) conjugado *de un número complejo* $a + bi$ *es el número complejo* $a - bi$. *Escribiremos el conjugado del número complejo* z *como* $\bar{z}$.

Ejemplo 3. Los conjugados de $-3 + 2i$, $4 - 7i$ y 6 son los siguientes:

$$\overline{-3 + 2i} = -3 - 2i, \qquad \overline{4 - 7i} = 4 + 7i,$$

y

$$\bar{6} = \overline{6 + 0i} = 6 - 0i = 6.$$

El siguiente resultado es una consecuencia inmediata de la definición de complejos conjugados.

Teorema D.2. *Un número complejo* z *es un número real si y sólo si* $z = \bar{z}$.

Para cualquier número complejo $z = a + bi$, $z\bar{z}$ es real y no negativo, ya que

$$z\bar{z} = (a + bi)(a - bi) = a^2 + b^2.$$

Este hecho puede utilizarse para definir el valor absoluto de un número complejo.

Definición. *El* valor absoluto *(o* módulo) *de un número complejo* $z = a + bi$ *es el número real* $\sqrt{a^2 + b^2}$. *Escribiremos el valor absoluto de* z *como* $|z|$. *Obsérvese que* $z\bar{z} = |z|^2$.

El hecho de que el producto de un número complejo por su conjugado sea real proporciona un método sencillo para determinar el cociente de dos números complejos, ya que si $c + di \neq 0$, entonces

$$\frac{a + bi}{c + di} = \frac{a + bi}{c + di} \cdot \frac{c - di}{c - di} = \frac{(ac + bd) + (bc - ad)i}{c^2 + d^2}$$

$$= \frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2}i.$$

Ejemplo 4. Ilustraremos el procedimiento descrito anteriormente calculando el cociente $(1 + 4i)/(3 - 2i)$:

$$\frac{1 + 4i}{3 - 2i} = \frac{1 + 4i}{3 - 2i} \cdot \frac{3 + 2i}{3 + 2i} = \frac{-5 + 14i}{9 + 4} = -\frac{5}{13} + \frac{14}{13}i.$$

El valor absoluto de un número complejo tiene las propiedades ordinarias del valor absoluto de un número real, tal como lo muestra el siguiente resultado.

Teorema D.3. *Sean* z *y* w *dos números complejos cualesquiera. Entonces*

(a) $|z + w| \leq |z| + |w|$.
(b) $|zw| = |z| \cdot |w|$.
(c) $|z| - |w| \leq |z + w|$.

DEMOSTRACIÓN. Sean $z = a + bi$ ỳ $w + c + di$, donde a, b, c y d son números reales.

(a) Obsérvese primero que

$$0 \leq (ad - bc)^2 = a^2d^2 - 2abcd + b^2c^2,$$

entonces $2abcd \leq a^2d^2 + b^2c^2$. Sumando $a^2c^2 + b^2d^2$ a ambos lados de la desigualdad se tendrá que

$$(ac + bd)^2 = a^2c^2 + 2abcd + b^2d^2$$
$$\leq a^2c^2 + a^2d^2 + b^2c^2 + b^2d^2 = (a^2 + b^2)(c^2 + d^2).$$

Tomando las raíces cuadradas, tenemos

$$ac + bd \leq \sqrt{a^2 + b^2}\sqrt{c^2 + d^2}.$$

Ahora bien,

$$\begin{aligned} |z + w|^2 &= |(a + c) + (b + d)i|^2 \\ &= (a + c)^2 + (b + d)^2 \\ &= a^2 + c^2 + b^2 + d^2 + 2(ac + bd) \\ &\leq a^2 + c^2 + b^2 + d^2 + 2\sqrt{a^2 + b^2}\sqrt{c^2 + d^2} \\ &= (\sqrt{a^2 + b^2} + \sqrt{c^2 + d^2})^2 \\ &= (|z| + |w|)^2. \end{aligned}$$

Tomando las raíces cuadradas, obtenemos (a).

(b) A partir de la definición de valor absoluto vemos que

$$\begin{aligned} |zw| &= |(a + bi)(c + di)| = |(ac - bd) + (bc + ad)i| \\ &= \sqrt{(ac - bd)^2 + (bc + ad)^2} = \sqrt{a^2c^2 + b^2d^2 + b^2c^2 + a^2d^2} \\ &= \sqrt{a^2 + b^2}\sqrt{c^2 + d^2} = |a + bi| \cdot |c + di| = |z| \cdot |w|. \end{aligned}$$

(c) Partiendo de (a) y (b) se tiene que

$$|z| = |(z + w) - w| \leq |z + w| + |-w| = |z + w| + |w|.$$

Entonces

$$|z| - |w| \leq |z + w|. \quad \blacksquare$$

Nuestra motivación para agrandar el conjunto de números reales al conjunto de números complejos fue la de obtener un campo tal que cada

polinomio de grado no nulo con coeficientes de ese campo tenga un cero. Nuestro siguiente resultado garantiza que el campo de números complejos tiene esta propiedad.

Teorema D.4. *(Teorema fundamental del álgebra.) Sean* $a_0, \ldots, a_n$ $(n \geq 1)$ *números complejos tales que* $a_n \neq 0$. *Entonces*

$$a_n z^n + a_{n-1} z^{n-1} + \ldots + a_1 z + a_0$$

tiene un cero en el campo de los números complejos.

Para la demostración consúltese: *Principles of Mathematical Analysis,* de Walter Rudin; McGraw-Hill Book Company, 1964.

El siguiente importante corolario se deriva del Teorema D.4 y del algoritmo de la división para polinomios (Teorema E.1).

Corolario. *Si* $p(z) = a_n z^n + \ldots + a_1 z + a_0$ *es un polinomio de grado* $n \geq 1$ *con coeficientes complejos, entonces existen números complejos* $c_1, \ldots,$ c_n *(no necesariamente distintos) tales que*

$$p(z) = a_n(z - c_1) \ldots (z - c_n).$$

Un campo se llama *algebraicamente cerrado* si tiene la propiedad de que todo polinomio con coeficientes de ese campo se descompone en un producto de factores de grado 1. Por lo tanto, el corolario anterior demuestra que el campo de números complejos es algebraicamente cerrado.

APENDICE E POLINOMIOS

En este Apéndice expondremos algunas propiedades básicas de los polinomios necesarias para los Capítulos 5 y 6. Para la definición de polinomios, véase la Sección 1.2.

Definición. *Un polinomio* $f(x)$ *divide a un polinomio* $g(x)$ *si existe un polinomio* $q(x)$ *tal que* $g(x) = f(x)q(x)$.

Nuestro primer resultado enseña que el largo proceso ordinario de división para polinomios con coeficientes reales es válido para polinomios con coeficientes de un campo cualquiera.

Teorema E.1. *(Algoritmo de la división para polinomios.) Sea* $f_1(x)$ *un polinomio de grado* n, *y sea* $f_2(x)$ *un polinomio de grado* $m \geq 0$. *Entonces, existen polinomios* $q(x)$ *y* $r(x)$ *tales que*

(a) *El grado de* $r(x)$ *es menor que* m.
(b) $f_1(x) = q(x)f_2(x) + r(x)$.
(c) $q(x)$ *y* $r(x)$ *son únicos con respecto a las condiciones* (a) *y* (b).

DEMOSTRACIÓN. Principiaremos estableciendo la existencia de $q(x)$ y $r(x)$ que satisfarán las condiciones (a) y (b). Si $n < m$, podemos tomar $q(x) = 0$ y $r(x) = f_1(x)$ que satisfacen (a) y (b).

Supóngase, por tanto, que $m \leq n$. En este caso estableceremos la existencia de $q(x)$ y $r(x)$ por inducción sobre n. Supóngase primero que $n = 0$; entonces $m \leq n$ implica que $m = 0$, de manera que $f_1(x)$ y $f_2(x)$ son constantes no nulas. Por lo tanto, podemos tomar a $q(x) = f_1(x)$ $(x)f_2^{-1}(x)$ y $r(x) = 0$ para satisfacer a (a) y (b).

Ahora supongamos que el teorema es cierto siempre que $f_1(x)$ tenga un grado menor que $n > 0$. Sea

$$f_1(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$$

y

$$f_2(x) = b_m x^m + b_{m-1}x^{m-1} + \ldots + b_1 x + b_0,$$

donde $m \leq n$. Defínase un polinomio $h(x)$ mediante

$$\begin{aligned} h(x) &= f_1(x) - a_n b_m^{-1} x^{n-m} f_2(x) \qquad (1) \\ &= (a_{n-1} - a_n b_m^{-1} b_{m-1})x^{n-1} + (a_{n-2} - a_n b_m^{-1} b_{m-2})x^{n-2} \\ &\qquad + \ldots + (a_0 - a_n b_m^{-1} b_0). \end{aligned}$$

Entonces $h(x)$ es un polinomio de grado menor que n. Consideraremos dos casos.

CASO 1. $h(x)$ es de grado menor que m. En este caso, sea $q(x) = a_n b_m^{-1} x^{n-m}$ y $r(x) = h(x)$. Entonces, de la Ecuación (1) se tiene

$$f_1(x) = q(x)f_2(x) + r(x),$$

y $r(x)$ tiene grado menor que m.

CASO 2. $h(x)$ es de grado mayor o igual que m. Como $h(x)$ tiene grado menor que n, podemos aplicar la hipótesis de inducción para obtener polinomios $q_1(x)$ y $r(x)$ tales que $r(x)$ es de grado menor que m y

$$h(x) = q_1(x)f_2(x) + r(x). \qquad (2)$$

Combinando las Ecuaciones (1) y (2) y resolviendo para $f_1(x)$, tenemos

$$f_1(x) = [a_n b_m^{-1} x^{n-m} + q_1(x)]f_2(x) + r(x).$$

En este caso sea $q(x) = a_n b_m^{-1} x^{n-m} + q_1(x)$, de manera que $f_1(x) = q$ $q(x)f_2(x) + r(x)$, donde $r(x)$ tiene grado menor que m. Esto demuestra la existencia de $q(x)$ y $r(x)$.

Demostraremos ahora la unicidad de q y r. Supóngase que $q_1(x)$, $q_2(x)$, $r_1(x)$ y $r_2(x)$ existen de modo que $r_1(x)$ y $r_2(x)$ tienen ambos un grado menor que m y

$$f_1(x) = q_1(x)f_2(x) + r_1(x) = q_2(x)f_2(x) + r_2(x).$$

Entonces

$$[q_1(x) - q_2(x)]f_2(x) = r_2(x) - r_1(x). \qquad (3)$$

El lado derecho de la Ecuación (3) es un polinomio de grado menor que m. Como $f_2(x)$ tiene grado m, debemos tener que $q_1(x) - q_2(x)$ es el polinomio cero. Por lo tanto, $q_1(x) = q_2(x)$, y por la Ecuación (3) $r_1(x) = r_2(x)$. ■

Dentro del contexto del Teorema E.1, llamamos a $q(x)$ y a $r(x)$, respectivamente, *cociente* y *residuo* de la división de $f_1(x)$ entre $f_2(x)$. Por ejemplo, el cociente y el residuo de la división del polinomio complejo

$$f_1(x) = (3 + i)x^5 - (1 - i)x^4 + 6x^3 + (-6 + 2i)x^2 + (2 + i)x + 1$$

entre el polinomio complejo

$$f_2(x) = (3 + i)x^2 - 2ix + 4$$

son

$$q(x) = x^3 + ix^2 - 2 \quad \text{y} \quad r(x) = (2 - 3i)x + 9.$$

Corolario 1. *Sea* $f(x)$ *un polinomio cuyo grado es al menos* 1, *y sea* $a \in F$. *Entonces* $f(a) = 0$ *si y sólo si* $x - a$ *divide a* $f(x)$.

DEMOSTRACIÓN. Supóngase que $x - a$ divide a $f(x)$. Entonces, existe un polinomio $q(x)$ tal que $f(x) = (x - a)q(x)$, por lo que $f(a) = (a - - a)q(a) = 0 \cdot q(a) = 0$.

Recíprocamente, supóngase que $f(a) = 0$. Por el Teorema E.1 existen polinomios $q(x)$ y $r(x)$ tales que $r(x)$ tiene grado menor que uno y

$$f(x) = q(x)(x - a) + r(x).$$

Sustituyendo a por x en la expresión anterior obtenemos $r(a) = 0$. Como $r(x)$ tiene grado menor que 1, debe ser el polinomio constante $r(x) = 0$. Luego, $f(x) = q(x)(x - a)$. ■

Para cualquier polinomio $f(x)$ con coeficientes de un campo F, un elemento $a \in F$ se llama *cero* de $f(x)$ si $f(a) = 0$. Con esta terminología, el corolario anterior establece que a es un cero de $f(x)$ si y sólo si $x - a$ divide a $f(x)$.

Corolario 2. *Cualquier polinomio de grado* $n \geq 1$ *tiene como máximo* n *ceros distintos.*

DEMOSTRACIÓN. La demostración se hará por inducción sobre n. El resultado es evidente si $n = 1$. Supóngase entonces que el resultado es cierto para algún entero positivo n, y sea $f(x)$ un polinomio de grado $n + 1$.

Si $f(x)$ no tiene ceros, no hay nada que demostrar. Por otra parte, si a es un cero de $f(x)$, por el Corolario 1 podemos escribir $f(x) = (x - a)q(x)$ para algún polinomio $q(x)$. Nótese que $q(x)$ debe ser de grado n; por lo tanto, por la hipótesis de inducción, $q(x)$ puede tener a lo más n ceros diferentes. Entonces, como cualquier cero de $f(x)$ diferente de a es también un cero de $q(x)$, $f(x)$ puede tener como máximo $n + 1$ ceros diferentes. ■

En el estudio de las formas canónicas surgen de manera natural polinomios que no tienen divisores comunes.

Definición. *Sean* f(x) *y* g(x) *polinomios, ambos de grado mayor que cero. Se dice que estos polinomios son* primos relativos *si no existe ningún polinomio de grado positivo que los divida a ambos.*

Por ejemplo, los polinomios $f(x) = x^2(x - 1)$ y $h(x) = (x - 1)(x - 2)(x - 1)(x - 2)$ no son primos relativos, puesto que $x - 1$ divide a $f(x)$ y a $h(x)$. Los polinomios $f(x)$ y $g(x) = (x - 2)(x - 3)$ son primos relativos, puesto que no tienen factores comunes de grado positivo.

El teorema siguiente establece que una combinación de polinomios primos relativos es igual al polinomio constante 1.

Teorema E.2. *Si* $f_1(x)$ *y* $f_2(x)$ *son polinomios primos relativos, existen polinomios* $q_1(x)$ *y* $q_2(x)$ *tales que* $q_1(x)f_1(x) + q_2(x)f_2(x) = 1$, *que es el polinomio constante de grado cero con valor* 1.

DEMOSTRACIÓN. Sin pérdida de generalidad, supóngase que el grado de $f_1(x)$ es mayor o igual que el grado de $f_2(x)$. Utilizaremos inducción matemática sobre el grado de $f_2(x)$. Supóngase que $f_2(x)$ tiene grado 1. Por el Teorema E.1 existen polinomios $q(x)$ y $r(x)$ tales que $r(x)$ tiene grado menor que 1 y que

$$f_1(x) = q(x)f_2(x) + r(x). \tag{4}$$

Nótese que $r(x)$ no puede ser el polinomio cero, puesto que $f_1(x)$ y $f_2(x)$ son primos relativos. Por tanto, $r(x)$ es una constante no nula c. Entonces la Ecuación (4) puede reescribirse como

$$(c^{-1})f_1(x) + (-c)^{-1}q(x)f_2(x) = 1. \tag{5}$$

Así, la conclusión vale con $q_1(x) = c^{-1}$ y $q_2(x) = (-c)^{-1}q(x)$. Ahora supóngase que el teorema se cumple cuando $f_2(x)$ tenga grado menor que n para algún entero $n \geq 2$ y supóngase que $f_2(x)$ tiene grado n. Por el Teorema E.1 existen polinomios $q(x)$ y $r(x)$ tales que $r(x)$ tenga grado menor que n y

$$f_1(x) = q(x)f_2(x) + r(x). \tag{6}$$

Como $f_1(x)$ y $f_2(x)$ son primos relativos, $r(x)$ no es el polinomio cero. Si $r(x)$ tiene grado cero, entonces $r(x)$ es una constante no nula c, y obtenemos la Ecuación (5) como antes. Supóngase entonces que $r(x)$ es de grado mayor que cero. Como $r(x)$ tiene grado menor que n, podemos aplicar la hipótesis de inducción a $f_2(x)$ y a $r(x)$, siempre que podamos demostrar que dichos polinomios son primos relativos. Supóngase lo contrario; entonces existe un polinomio no nulo $g(x)$ que divide a $f_2(x)$ y a $r(x)$. Luego, existen polinomios $h_1(x)$ y $h_2(x)$ tales que

$$r(x) = g(x)h_1(x) \quad \text{y} \quad f_2(x) = g(x)h_2(x). \tag{7}$$

Combinando las Ecuaciones (6) y (7), obtenemos

$$f_1(x) = [q(x)h_2(x) + h_1(x)]g(x),$$

y entonces $g(x)$ divide a $f_1(x)$. Pero $g(x)$ divide a $f_2(x)$, contradiciendo el hecho de que $f_1(x)$ y $f_2(x)$ son primos relativos, de modo que $r(x)$ y $f_2(x)$ son primos relativos. Por tanto, por la hipótesis de inducción, existen $g_1(x)$ y $g_2(x)$ tales que

$$g_1(x)f_2(x) + g_2(x)r(x) = 1. \tag{8}$$

Combinando las Ecuaciones (6) y (8), tenemos

$$g_1(x)f_2(x) + g_2(x)[f_1(x) - q(x)f_2(x)] = 1.$$

De donde

$$g_2(x)f_1(x) + [g_1(x) - g_2(x)q(x)]f_2(x) = 1.$$

Haciendo $q_1(x) = g_2(x)$ y $q_2(x) = g_1(x) - g_2(x)q(x)$ obtenemos la conclusión deseada. ■

Ejemplo 1. Para los polinomios primos relativos $f_1(x) = x^2(x-1)$ y $f_2(x) = (x-2)(x-3)$, se verifica fácilmente que

$$q_1(x)f_1(x) + q_2(x)f_2(x) = 1,$$

donde

$$q_1(x) = \tfrac{1}{36}(-7x + 23) \quad \text{y} \quad q_2(x) = \tfrac{1}{36}(x^2 + 5x + 6).$$

A lo largo de los Capítulos 5, 6 y 7 consideramos operadores lineales que son polinomios en algún operador particular T y matrices que son polinomios en una matriz particular A. Para estos operadores y matrices es conveniente la siguiente notación.

Definiciones. *Sea*

$$f(x) = a_0 + a_1x + \dots + a_nx^n$$

un polinomio con coeficientes de un campo F. *Si* T *es un operador lineal en* V, *un espacio vectorial sobre* F, *definimos* f(T) *mediante*

$$f(T) = a_0I + a_1T + \dots + a_nT^n.$$

De manera análoga, si A *es una matriz de* $n \times n$ *con elementos de* F, *definimos* f(A) *mediante*

$$f(A) = a_0I + a_1A + \ldots + a_nA^n.$$

Ejemplo 2. Sea T un operador lineal en R^2 definido mediante $T(a, b) = (2a + b, a - b)$ y sea $f(x) = x^2 + 2x - 3$. Como $T^2(a, b) = (5a + b, a + 2b)$,

$$\begin{aligned} f(T)(a, b) &= (T^2 + 2T - 3I)(a, b) \\ &= (5a + b, a + 2b) + (4a + 2b, 2a - 2b) - 3(a, b) \\ &= (6a + 3b, 3a - 3b). \end{aligned}$$

Análogamente, si

$$A = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix},$$

entonces

$$\begin{aligned} f(A) = A^2 + 2A - 3I &= \begin{pmatrix} 5 & 1 \\ 1 & 2 \end{pmatrix} + 2\begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} - 3\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 6 & 3 \\ 3 & -3 \end{pmatrix}. \end{aligned}$$

Los siguientes tres teoremas utilizan esta notación.

Teorema E.3. *Sea* f(x) *un polinomio con coeficientes de un campo* F, *y sea* T *un operador lineal en* V, *donde* V *es un espacio vectorial sobre* F. *Entonces*

(a) f(T) *es un operador lineal en* V.
(b) *Si* β *es una base ordenada finita para* V *y* $A = [T]_\beta$, *entonces* $[f(T)]_\beta = f(A)$.

DEMOSTRACIÓN. Ejercicio.

Teorema E.4. *Sea* T *un operador lineal en un espacio vectorial* V *sobre* F, *y sea* A *una matriz cuadrada con elementos de* F. *Entonces, para polinomios cualesquiera* $f_1(x)$ *y* $f_2(x)$, *con coeficientes de* F

(a) $f_1(T)f_2(T) = f_2(T)f_1(T)$.
(b) $f_1(A)f_2(A) = f_2(A)f_1(A)$.

DEMOSTRACIÓN. Ejercicio.

Teorema E.5. *Sea* T *un operador lineal sobre un espacio vectorial* V *sobre un campo* F, *y sea* A *una matriz de* $n \times n$ *con elementos de* F. *Si* $f_1(x)$ *y*

$f_2(x)$ *son polinomios primos relativos con elementos de* F, *entonces existen polinomios* $q_1(x)$ *y* $q_2(x)$ *con elementos de* F *tales que*

(a) $q_1(T)f_1(T) + q_2(T)f_2(T) = I$.
(b) $q_1(A)f_1(A) + q_2(A)f_2(A) = I$.

DEMOSTRACIÓN. Ejercicio.

En los Capítulos 5 y 6 nos interesa determinar cuándo un operador lineal T en un espacio vectorial dimensionalmente finito puede "diagonalizarse", así como encontrar una representación sencilla (canónica) de T. Ambos problemas son afectados por la factorización de un cierto polinomio determinando por T (el "polinomio característico" de T). En este asunto tienen un papel importante algunas clases de polinomios.

Definiciones. *Un polinomio* f(x) *con coeficientes de un campo* F, *se llama* mónico *si su coeficiente principal (el de la potencia mayor) es* 1. *Si* f(x) *tiene grado positivo y no puede expresarse como un producto de polinomios con coeficientes de* F, *cada uno de grado positivo,* f(x) *se denomina* irreducible.

Obsérvese que, el que un polinomio sea o no irreducible, depende del campo del que provengan sus coeficientes. Por ejemplo, $f(x) = x^2 + 1$ es irreducible en el campo de los números reales, pero no lo es en el campo de los números complejos, puesto que $x^2 + 1 = (x + i)(x - i)$.

Evidentemente, un polinomio de grado 1 es irreducible. Más aún, para polinomios con coeficientes de un campo algebraicamente cerrado, los polinomios de grado 1 son los únicos polinomios irreducibles.

Pueden establecerse fácilmente los siguientes hechos.

Teorema E.6. *Sean* $\phi(x)$ *y* f(x) *polinomios con coeficientes de un campo* F. *Si* $\phi(x)$ *es irreducible y* $\phi(x)$ *no divide a* f(x), *entonces* $\phi(x)$ *y* f(x) *son primos relativos.*

DEMOSTRACIÓN. Ejercicio.

Teorema E.7. *Cualquier par de polinomios mónicos irreducibles distintos son primos relativos.*

DEMOSTRACIÓN. Ejercicio.

Ahora estableceremos un resultado que conducirá a la demostración del teorema de factorización única para polinomios, el cual establece que todo polinomio de grado positivo es expresable de manera única como un producto de polinomios irreducibles mónicos multiplicado por una constante.

Teorema E.8. *Sean* $f(x)$, $g(x)$ *y* $\phi(x)$ *polinomios con coeficientes del mismo campo. Si* $\phi(x)$ *es irreducible y divide al producto* $f(x)g(x)$, *entonces* $\phi(x)$ *divide a* $f(x)$ *o* $\phi(x)$ *divide a* $g(x)$.

DEMOSTRACIÓN. Supóngase que $\phi(x)$ no divide a $f(x)$. Entonces $\phi(x)$ y $f(x)$ son primos relativos por el Teorema E.6 y, por lo tanto, existen polinomios $q_1(x)$ y $q_2(x)$ tales que

$$1 = q_1(x)\phi(x) + q_2(x)f(x).$$

Al multiplicar ambos lados de la ecuación por $g(x)$ se tiene

$$g(x) = q_1(x)\phi(x)g(x) + q_2(x)f(x)g(x). \tag{9}$$

Como $\phi(x)$ divide a $f(x)g(x)$, existe un polinomio $h(x)$ tal que $f(x)g(x) = \phi(x)h(x)$. Y la Ecuación (9) se convierte en

$$g(x) = q_1(x)\phi(x)g(x) + q_2(x)\phi(x)h(x) = \\ = \phi(x)[q_1(x)g(x) + q_2(x)h(x)].$$

Por lo que $\phi(x)$ divide a $g(x)$. ■

Corolario. *Sean* $\phi(x)$, $\phi_1(x)$, $\phi_2(x)$, $\ldots$, $\phi_n(x)$ *polinomios mónicos irreducibles con coeficientes del mismo campo. Si* $\phi(x)$ *divide al producto* $\phi_1(x)\phi_2(x)\ldots\phi_n(x)$, *entonces* $\phi(x) = \phi_i(x)$ *para alguna* $i(i = 1, 2, \ldots, n)$.

DEMOSTRACIÓN. Demostraremos el corolario por inducción sobre n. Para $n = 1$ el resultado es una consecuencia inmediata del Teorema E.7. Supóngase entonces que el corolario es cierto para cualesquier $n - 1$ polinomios mónicos irreducibles y que contamos con n polinomios mónicos irreducibles $\phi_1(x)$, $\phi_2(x)$, $\ldots$, $\phi_n(x)$. Si $\phi(x)$ divide al producto

$$\phi_1(x)\phi_2(x)\ldots\phi_n(x) = [\phi_1(x)\phi_2(x)\ldots\phi_{n-1}(x)]\phi_n(x),$$

entonces, por el Teorema E.8, $\phi(x)$ divide al producto $\phi_1(x)\phi_2(x)\ldots$ $\ldots\phi_{n-1}(x)$ o $\phi(x)$ divide a $\phi_n(x)$. En el primer caso $\phi(x) = \phi_i(x)$ para alguna $i(i = 1, 2, \ldots, n - 1)$, por la hipótesis de inducción; en el segundo caso, por el Teorema E.7, $\phi(x) = \phi_n(x)$. ■

Ahora ya somos capaces de establecer el teorema de factorización única, que se utiliza a lo largo de los Capítulos 5 y 6.

Teorema E.9. *(Teorema de factorización única para polinomios.) Para cualquier polinomio* $f(x)$ *de grado positivo, existen una constante única* c, *polinomios mónicos irreducibles* $\phi_1(x)$, $\phi_2(x)$, $\ldots$, $\phi_k(x)$, *y enteros positivos únicos* n_1, n_2, $\ldots$, n_k *tales que*

$$f(x) = c[\phi_1(x)]^{n_1}[\phi_2(x)]^{n_2}\ldots[\phi_k(x)]^{n_k}.$$

DEMOSTRACIÓN. Principiaremos demostrando la existencia de tal factorización utilizando inducción sobre el grado de $f(x)$. Si $f(x)$ es de gra-

do 1, entonces $f(x) = ax + b$ para algunas constantes a y b con $a \neq 0$. Haciendo $\phi(x) = x + b/a$, tenemos que $f(x) = a\phi(x)$. Como $\phi(x)$ es un polinomio mónico irreducible queda demostrado el resultado en este caso. Ahora supongamos que la conclusión es cierta para cualquier polinomio de grado positivo menor que algún entero $n > 1$ y sea $f(x)$ un polinomio de grado n. Entonces

$$f(x) = a_n x^n + \ldots + a_1 x + a_0$$

para algunos escalares a_i con $a_n \neq 0$. Si $f(x)$ es irreducible, entonces

$$f(x) = a_n \left(x^n + \frac{a_{n-1}}{a_n} x^{n-1} + \ldots + \frac{a_1}{a_n} x + \frac{a_0}{a_n} \right).$$

es una representación de $f(x)$ como un producto de a_n por un polinomio mónico irreducible. Si $f(x)$ no es irreducible, entonces $f(x) = g(x)h(x)$ para algunos polinomios $g(x)$ y $h(x)$, cada uno de grado positivo menor que n. La hipótesis de inducción garantiza que $g(x)$ y $h(x)$ se factorizan como productos de una constante por potencias de diferentes polinomios mónicos irreducibles. Por lo tanto, $f(x) = g(x)h(x)$ también se factoriza de este modo. Así pues, en ambos casos $f(x)$ puede ser factorizado como un producto de una constante por potencias de polinomios mónicos irreducibles.

Falta establecer la unicidad de tal factorización. Supóngase que

$$\begin{aligned} f(x) &= c[\phi_1(x)]^{n_1}[\phi_2(x)]^{n_2} \ldots [\phi_k(x)]^{n_k} \\ &= d[\psi_1(x)]^{m_1}[\psi_2(x)]^{m_2} \ldots [\psi_r(x)]^{m_r}, \end{aligned} \tag{10}$$

donde c y d son constantes, $\phi_i(x)$ y $\psi_j(x)$ son polinomios mónicos irreducibles, y n_i y m_j son enteros positivos ($i = 1, 2, \ldots, k$ y $j = 1, 2, \ldots, r$). Claramente se ve que tanto c como d deben ser el coeficiente principal de $f(x)$; por lo tanto, $c = d$, y la Ecuación (10) se transforma en

$$[\phi_1(x)]^{n_1}[\phi_2(x)]^{n_2} \ldots [\phi_k(x)]^{n_k} = [\psi_1(x)]^{m_1}[\psi_2(x)]^{m_2} \ldots [\psi_r(x)]^{m_r}. \tag{11}$$

Así tenemos que $\phi_i(x)$ divide al lado derecho de la Ecuación (11) para $i = 1, 2, \ldots, k$. Consecuentemente, por el corolario del Teorema E.8, para cada $i (i = 1, 2, \ldots, k)$ $\phi_i(x) = \psi_j(x)$ para alguna $j = 1, 2, \ldots, r$, y para cualquier $j (j = 1, 2, \ldots, r)$ $\psi_j(x) = \phi_i(x)$ para alguna $i = 1, 2, \ldots, k$. Concluimos que $r = k$ y que, renumerando en caso de ser necesario, $\phi_i(x) = \psi_i(x)$ para $i = 1, 2, \ldots, k$. Supóngase que $n_i \neq m_i$ para alguna i. Sin perder generalidad podemos suponer que $i = 1$ y $n_1 > m_1$. Entonces, eliminando $[\phi_1(x)]^{m_1}$ de ambos lados de la Ecuación (11) tenemos

$$[\phi_1(x)]^{n_1-m_1}[\phi_2(x)]^{n_2} \ldots [\phi_k(x)]^{n_k} = [\phi_2(x)]^{m_2} \ldots [\phi_k(x)]^{m_k}. \tag{12}$$

Como $n_1 - m_1 > 0$, $\phi_1(x)$ divide a la parte izquierda de la Ecuación (12) y por tanto también divide al lado derecho. Así, $\phi_1(x) = \phi_i(x)$ para alguna $i = 2, \ldots, k$, por el corolario al Teorema E.8. Pero esto contradice

el hecho de que $\phi_1(x)$, $\phi_2(x)$, ..., $\phi_k(x)$ son distintos. Por lo tanto, las factorizaciones de $f(x)$ en la Ecuación (10) son las mismas. ■

Es a menudo útil considerar un polinomio $f(x) = a_n x^n + \ldots + a_1 x + a_0$ con coeficientes del campo F como una función f: $F \to F$. En este caso, el valor de f en $c \in F$ es $f(c) = a_n c^n + \ldots + a_1 c + a_0$. Desafortunadamente, para campos cualesquiera F, no existe correspondencia uno-a-uno entre polinomios y funciones polinomiales. Por ejemplo, si $f(x) = x^2$ y $g(x) = x$ son dos polinomios del campo Z_2 (como se definió en el Ejemplo 4 del Apéndice C), entonces $f(x)$ y $g(x)$ tienen grados diferentes y, por lo tanto, no son iguales como polinomios. Pero $f(a) = g(a)$ para toda $a \in Z_2$, y así f y g son funciones polinomiales iguales. Nuestro resultado final muestra que esta anomalía no puede ocurrir si F es un campo infinito.

Teorema E.10. *Sean* f(x) *y* g(x) *polinomios con coeficientes de un campo infinito* F. *Si* f(a) = g(a) *para toda* a ∈ F, *entonces* f(x) *y* g(x) *son iguales.*

DEMOSTRACIÓN. Supóngase que $f(a) = g(a)$ para toda $a \in F$. Defínase $h(x) = f(x) - g(x)$ y supóngase que $h(x)$ es de grado $n \geq 1$. Se tiene del corolario al Teorema E.9 que $h(x)$ puede tener como máximo n ceros. Pero $h(a) = f(a) - g(a) = 0$ para cualquier $a \in F$ contradiciendo la hipótesis de que $h(x)$ tiene grado positivo. Así, $h(x)$ es un polinomio constante y como $h(a) = 0$ para cada $a \in F$ se tiene que $h(x)$ es el polinomio cero. Por lo tanto, $f(x) = g(x)$. ■

Respuestas a los ejercicios seleccionados

SECCION 1.1

1. Sólo los pares de los incisos (b) y (c) son paralelos.

2. (a) $(3, -2, 4) + t(-8, 9, -3)$
(c) $(3, 7, 2) + t(0, 0, -10)$

3. (a) $(2, -5, -1) + t_1(-2, 9, 7) + t_2(-5, 12, 2)$
(c) $(-8, 2, 0) + t_1(9, 1, 0) + t_2(14, -7, 0)$

SECCION 1.2

1. (a) V (b) F (c) F (d) F (e) V (f) F
(g) F (h) F (i) V (j) V (k) V

3. $M_{13} = 3$, $M_{21} = 4$, y $M_{22} = 5$

4. (a) $\begin{pmatrix} 6 & 3 & 2 \\ -4 & 3 & 9 \end{pmatrix}$ (c) $\begin{pmatrix} 8 & 20 & -12 \\ 4 & 0 & 28 \end{pmatrix}$
(e) $2x^4 + x^3 + 2x^2 - 2x + 10$
(g) $10x^7 - 30x^4 + 40x^2 - 15x$

13. No, (VS 4) falla.

14. Sí.

15. No.

SECCION 1.3

1. (a) F (b) F (c) V (d) F (e) V (f) F

2. (a) $\begin{pmatrix} -4 & 5 \\ 2 & -1 \end{pmatrix}$; la traza es -5.

(c) $\begin{pmatrix} -3 & 0 & 6 \\ 9 & -2 & 1 \end{pmatrix}$ (e) $\begin{pmatrix} 1 \\ -1 \\ 3 \\ 5 \end{pmatrix}$

11. No, el conjunto no es cerrado bajo la adición.
14. Sí.

SECCION 1.4

1. (a) V (b) F (c) V (d) F (e) V (f) F
2. (a) $\{x_2(1, 1, 0, 0) + x_4(-3, 0, -2, 1) + (5, 0, 4, 0): x_2, x_4 \in R\}$
(c) No hay soluciones.
(e) $\{x_3(10, -3, 1, 0, 0) + x_4(-3, 2, 0, 1, 0) + (-4, 3, 0, 0, 5): x_3, x_4 \in R\}$
3. (a) Sí. (c) No. (e) No.
4. (a) Sí. (c) Sí. (e) No.

SECCION 1.5

1. (a) F (b) V (c) F (d) F (e) V (f) V
5. $\left\{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\right\}$

SECCION 1.6

1. (a) F (b) V (c) F (d) F (e) V (f) F
(g) F (h) V (i) F (j) V (k) V (l) F
2. (a) Sí. (c) Sí. (e) No.
3. (a) No. (c) No. (e) No.
4. No.
5. No.
8. $\{x_1, x_3, x_5, x_7\}$
9. $(a_1, a_2, a_3, a_4) = a_1x_1 + (a_2 - a_1)x_2 + (a_3 - a_2)x_3 + (a_4 - a_3)x_4$
10. $\dim(W_1) = 3$, $\dim(W_2) = 2$, $\dim(W_1 + W_2) = 4$, y $\dim(W_1 \cap W_2) = 1$
17. $n^2 - 1$
19. $\frac{1}{2}n(n - 1)$

SECCION 1.7

1. (a) F (b) F (c) F (d) V (e) V (f) V

SECCION 2.1

1. (a) V (b) F (c) F (d) V (e) V (f) F
   (g) F (h) V (i) F
2. La nulidad es 1, y el rango es 2. T no es uno a uno pero es sobreyectiva
4. La nulidad es 4, y el rango es 2. T no es ni uno a uno ni sobreyectiva
5. La nulidad es 0, y el rango es 3. T es uno a uno pero no sobreyectiva
10. T(2, 3) = (5, 11). T es uno a uno
12. No.

SECCION 2.2

1. (a) V (b) V (c) F (d) V (e) V (f) F
2. (a) $\begin{pmatrix} 2 & -1 \\ 3 & 4 \\ 1 & 0 \end{pmatrix}$
2.

3. $[T]_\beta^\gamma = \begin{pmatrix} -\frac{1}{3} & -1 \\ 0 & 1 \\ \frac{2}{3} & 0 \end{pmatrix}$ y $[T]_\alpha^\gamma = \begin{pmatrix} -\frac{7}{3} & -\frac{11}{3} \\ 2 & 3 \\ \frac{2}{3} & \frac{4}{3} \end{pmatrix}$

5. (a) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ (b) $\begin{pmatrix} 0 & 1 & 0 \\ 2 & 2 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ (e) $\begin{pmatrix} 1 \\ -2 \\ 0 \\ 4 \end{pmatrix}$

9. $\begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$

SECCION 2.3

1. (a) F (b) V (c) F (d) V′ (e) F (f) F
   (g) F (h) F (i) V (j) V
2. $A(2B+3C) = \begin{pmatrix} 20 & -9 & 18 \\ 5 & 10 & 8 \end{pmatrix}$ y $A(BD) = \begin{pmatrix} 29 \\ -26 \end{pmatrix}$

3. $[\mathsf{T}]_\beta = \begin{pmatrix} 2 & 3 & 0 \\ 0 & 3 & 6 \\ 0 & 0 & 4 \end{pmatrix}$, $[\mathsf{U}]_\beta^\gamma = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$, y $[\mathsf{UT}]_\beta^\gamma = \begin{pmatrix} 2 & 6 & 6 \\ 0 & 0 & 4 \\ 2 & 0 & -6 \end{pmatrix}$

4. (a) $\begin{pmatrix} 1 \\ -1 \\ 4 \\ 6 \end{pmatrix}$ (c) (5)

11. (a) No. (b) No.

SECCION 2.4

1. (a) F (b) V (c) F (d) F (e) V (f) F
(g) V (h) V (i) V

17. (b) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

SECCION 2.5

1. (a) F (b) V (c) V (d) F (e) V

2. (a) $\begin{pmatrix} a_1 & a_1 \\ a_2 & b_2 \end{pmatrix}$ (c) $\begin{pmatrix} -1 & 3 \\ -2 & 5 \end{pmatrix}$

3. (a) $\begin{pmatrix} a_2 & b_2 & c_2 \\ a_1 & b_1 & c_1 \\ a_0 & b_0 & c_0 \end{pmatrix}$ (c) $\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ -3 & 2 & 1 \end{pmatrix}$ (e) $\begin{pmatrix} 5 & -6 & 3 \\ 0 & 4 & -1 \\ 3 & -1 & 2 \end{pmatrix}$

4. (a) $A = \begin{pmatrix} 3 & -1 \\ 2 & 4 \\ -1 & 1 \end{pmatrix}$ y $B = \begin{pmatrix} 5 & -4 \\ 7 & -2 \\ -1 & 2 \end{pmatrix}$

(b) $Q = \begin{pmatrix} 2 & -1 \\ 1 & 1 \end{pmatrix}$ y $P = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix}$

SECCION 2.6

1. (a) F (b) V (c) V (d) V (e) F (f) V
(g) V (h) F

2. Las funciones de los incisos (a), (c), (e) y (f) son funciones lineales.

3. (a) $\mathsf{f}_1(x, y, z) = x - \frac{1}{2}y$, $\mathsf{f}_2(x, y, z) = \frac{1}{2}y$, y $\mathsf{f}_3(x, y, z) = -x + z$

5. La base para V es $\{p_1(x), p_2(x)\}$, donde $p_1(x) = 2 - 2x$ y $p_2(x) = -\frac{1}{2} + x$.

7. (a) $T^t(f) = g$, donde $g(a + bx) = -3a - 4b$

(b) $[T^t]_{\gamma*}^{\beta*} = \begin{pmatrix} -1 & 1 \\ -2 & 1 \end{pmatrix}$ (c) $[T]_\beta^\gamma = \begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}$

SECCION 2.7

1. (a) V (b) V (c) F (d) F (e) V (f) F
(g) V

2. (a) F (b) V (c) V (d) V (e) F

3. (a) $\{e^{-t}, te^{-t}\}$ (c) $\{e^{-t}, te^{-t}, e^t, te^t\}$
(e) $\{e^{-t}, e^t \cos 2t, e^t \operatorname{sen} 2t\}$

4. (a) $\{e^{(1+\sqrt{5})t/2}, e^{(1-\sqrt{5})t/2}\}$ (c) $\{1, e^{-4t}, e^{-2t}\}$

SECCION 3.1

1. (a) V (b) F (c) V (d) F (e) V (f) F
(g) V (h) F (i) V

2. Añadiendo -2 veces la columna 1 a la columna 2 transforma A en B.

SECCION 3.2

1. (a) F (b) F (c) V (d) V (e) F (f) V
(g) V (h) V (i) V

2. (a) 2 (c) 2 (e) 3 (g) 1

4. (a) $D = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$; el rango es 2.

5. (a) El rango es 2, y la inversa es $\begin{pmatrix} -1 & -2 \\ 1 & -1 \end{pmatrix}$

(c) El rango es 3 y la inversa es $\begin{pmatrix} \frac{1}{6} & -\frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ -\frac{1}{6} & \frac{1}{3} & \frac{1}{2} \end{pmatrix}$

(e) El rango es 3; por lo que no existe inversa.

6. (a) $T^{-1}(ax^2 + bx + c) = -ax^2 - (4a + b)x - (10a + 2b + c)$

(c) $T^{-1}(ax^2 + bx + c) = (a, -\frac{1}{2}b + \frac{1}{2}c, -a + \frac{1}{2}b + \frac{1}{2}c)$

7. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

SECCION 3.3

1. (a) F (b) F (c) V (d) F (e) F (f) F
(g) V (h) F

2. (a) $\left\{\begin{pmatrix}1\\2\\3\end{pmatrix}\right\}$ (c) $\left\{\begin{pmatrix}2\\-1\\0\\0\end{pmatrix}, \begin{pmatrix}3\\0\\1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\0\\-1\end{pmatrix}\right\}$

3. (a) $\left\{\begin{pmatrix}1\\1\\1\end{pmatrix} + t\begin{pmatrix}1\\2\\3\end{pmatrix}: \quad t \in R\right\}$

(c) $\left\{\begin{pmatrix}1\\0\\0\\0\end{pmatrix} + t_1\begin{pmatrix}2\\-1\\0\\0\end{pmatrix} + t_2\begin{pmatrix}3\\0\\1\\0\end{pmatrix} + t_3\begin{pmatrix}1\\0\\0\\-1\end{pmatrix}: \quad t_1, t_2, t_3 \in R\right\}$

4. (b) $A^{-1} = \begin{pmatrix}\frac{1}{3} & 0 & \frac{1}{3}\\ \frac{1}{9} & \frac{1}{3} & -\frac{2}{9}\\ -\frac{4}{9} & \frac{2}{3} & -\frac{1}{9}\end{pmatrix}$ (c) $\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \begin{pmatrix}3\\0\\-2\end{pmatrix}$

6. $\mathrm{T}^{-1}\{(1, 11)\} = \left\{\begin{pmatrix}\frac{11}{2}\\-\frac{9}{2}\\0\end{pmatrix} + t\begin{pmatrix}1\\-1\\2\end{pmatrix}: \quad t \in R\right\}$

7. Los sistemas de los incisos (b), (c) y (d) tienen soluciones.

10. El agricultor, el sastre y el carpintero deben de tener ingresos en las proporciones 4: 3: 4.

11. Deben de tener 7.8 unidades del primer bien y 9.5 unidades del segundo.

SECCION 3.4

1. (a) F (b) V (c) V (d) V (e) F (f) V
(g) V

4. (a) $\left\{\begin{pmatrix}\frac{4}{3}\\\frac{1}{3}\\0\\0\end{pmatrix} + t\begin{pmatrix}1\\-1\\1\\2\end{pmatrix}: \quad t \in R\right\}$

(c) No hay soluciones.

SECCION 4.1

1. (a) V (b) F (c) V (d) F (e) V (f) F

2. (a) 30 (c) −8

3. (a) $-10 + 15i$ (c) -24
4. (a) 19 (c) 14

SECCION 4.2

1. (a) V (b) V (c) F (d) F (e) V (f) F
3. (a) -34 (c) -49
4. Las funciones en los incisos (c), (d) y (g) son 3− lineales.

SECCION 4.3

1. (a) V (b) V (c) F (d) V (e) F (f) V
(g) F (h) V (i) V (j) F (k) V
2. (a) 90 (c) 0
3. (a) 100 (c) 0 (e) 86 (g) $-180 + 40i$

SECCION 4.4

1. (a) F (b) F (c) F (d) F
2. (a) $\begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix}$ (c) $\begin{pmatrix} -3i & 0 & 0 \\ 4 & -1+i & 0 \\ 10+16i & -5-3i & 3+3i \end{pmatrix}$

(e) $\begin{pmatrix} 6 & 22 & 12 \\ 12 & -2 & 24 \\ 21 & -38 & -27 \end{pmatrix}$ (g) $\begin{pmatrix} 18 & 28 & -6 \\ -20 & -21 & 37 \\ 48 & 14 & -16 \end{pmatrix}$

3. (a) $x_1 = \dfrac{b_1a_{22} - b_2a_{12}}{a_{11}a_{22} - a_{12}a_{21}}, \quad x_2 = \dfrac{b_2a_{11} - b_1a_{21}}{a_{11}a_{22} - a_{12}a_{21}}$
(c) $x_1 = -1,\ x_2 = -1.2,\ x_3 = -1.4$
(e) $x_1 = -43,\ x_2 = -109,\ x_3 = -17$

SECCION 4.5

1. (a) V (b) V (c) F (d) F (e) V (f) V
2. (a) -1 (c) $(-1)^{n(n+1)/2}$
(e) $(-1)^{n-1}$

SECCION 4.6

1. (a) V (b) V (c) V (d) F (e) F (f) V
(g) V (h) F (i) V (j) V (k) V

2. (a) 22 (c) $2 - 4i$
3. (a) -12 (c) 22 (e) -3
4. (a) 88 (c) -6 (e) $17 - 3i$ (g) $24 + 24i$

SECCION 5.1

1. (a) F (b) V (c) V (d) F (e) F (f) F
(g) F (h) V (i) V (j) F (k) F

2. $Q = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ $[L_A]_\beta = \begin{pmatrix} 6 & 11 \\ -2 & -4 \end{pmatrix}$

3. (a) Los eigenvalores son 4 y -1, y una base de eigenvectores es

$$\left\{ \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\} \qquad Q = \begin{pmatrix} 2 & 1 \\ 3 & -1 \end{pmatrix}.$$

(c) Los eigenvalores son 1 y -1, y una base de eigenvectores es

$$\left\{ \begin{pmatrix} 1 \\ 1 - i \end{pmatrix}, \begin{pmatrix} 1 \\ -1 - i \end{pmatrix} \right\}. \qquad Q = \begin{pmatrix} 1 & 1 \\ 1 - i & -1 - 1 \end{pmatrix}.$$

4. Los eigenvalores son 1, 2 y 3, y una base de eigenvectores es $\{1, x, x^2\}$.

SECCION 5.2

1. (a) F (b) F (c) V (d) F (e) F (f) V
(g) V (h) F (i) V

2. (a) No diagonalizable. (c) $Q = \begin{pmatrix} 1 & 4 \\ 1 & -3 \end{pmatrix}$

(e) No diagonalizable. (g) $Q = \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$

3. (a) No diagonalizable. (c) No diagonalizable.

9.

$$A^n = \begin{pmatrix} \dfrac{5^n}{3} + \dfrac{2(-1)^n}{3} & \dfrac{2(5^n)}{3} - \dfrac{2(-1)^n}{3} \\ \dfrac{5^n}{3} - \dfrac{(-1)^n}{3} & \dfrac{2(5^n)}{3} + \dfrac{(-1)^n}{3} \end{pmatrix}$$

16. $X(t) = c_1 e^{3t} \begin{pmatrix} -2 \\ 1 \end{pmatrix} + c_2 e^{-2t} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

SECCION 5.3

1. (a) V (b) V (c) F (d) F (e) V (f) V
(g) V (h) F (i) F (j) V

2. (a) $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ (c) $\begin{pmatrix} \frac{7}{13} & \frac{7}{13} \\ \frac{6}{13} & \frac{6}{13} \end{pmatrix}$ (e) No existe límite.

(g) $\begin{pmatrix} -1 & 0 & -1 \\ -4 & 1 & -2 \\ 2 & 0 & 2 \end{pmatrix}$ (i) No existe límite.

6. Un mes después de su llegada 25% de los pacientes se han recuperado, 20% son ambulatorios, 41% están en cama y 14% han muerto; eventualmente $\frac{59}{90}$ se recuperan y $\frac{31}{90}$ mueren.

7. $\frac{4}{7}$

8. Sólo las matrices de los incisos (a) y (b) son matrices regulares de transición.

9. (a) $\begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$ (c) No existen límites.

(e) $\begin{pmatrix} 0 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & 0 & 1 \end{pmatrix}$ (g) $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 1 \end{pmatrix}$

10. (a) $\begin{pmatrix} 0.225 \\ 0.441 \\ 0.334 \end{pmatrix}$ después de dos etapas y eventualmente $\begin{pmatrix} 0.20 \\ 0.60 \\ 0.20 \end{pmatrix}$

(c) $\begin{pmatrix} 0.368 \\ 0.350 \\ 0.282 \end{pmatrix}$ después de dos etapas y eventualmente $\begin{pmatrix} 0.50 \\ 0.20 \\ 0.30 \end{pmatrix}$

(e) $\begin{pmatrix} 0.329 \\ 0.334 \\ 0.337 \end{pmatrix}$ después de dos etapas y eventualmente $\begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}$

12. $\frac{9}{19}$ nuevas, $\frac{6}{19}$ utilizadas una vez y $\frac{4}{19}$ utilizadas dos veces.

13. En 1985 24% poseerán autos grandes, 34% poseerán autos de tamaño mediano y 42% poseerán autos pequeños; las proporciones en un tiempo cualquiera son 0.10, 0.30 y 0.60.

18. $e^0 = I$ y $e^I = eI$.

SECCION 5.4

1. (a) F (b) V (c) V
2. Los subespacios de los incisos (a), (c) y (d) son T-invariantes.

SECCION 5.5

1. (a) F (b) F (c) V (d) V (e) V
2. (a) $\{e_1, e_2, e_3, e_4\}$ (c) $\left\{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right\}$
3. (a) (i) $-t(1-t)(t^2-3t+3)$ (ii) $-t(1-t)(t^2-3t+3)$
   (c) (i) $1-t$ (ii) $(t-1)^3(t+1)$

SECCION 5.6

1. (a) F (b) V (c) F (d) F (e) V (f) F
   (g) F (h) V
2. (a) $(t-1)(t-3)$ (c) $(t-1)^2(t-2)$
3. (a) $(t-2)^3$ (c) $(t-1)(t+1)$
5. Los operadores diagonalizables en R^2 que satisfacen a $\mathsf{T}^3 - 2\mathsf{T}^2 + \mathsf{T} = \mathsf{T}_0$ son T_0, I y aquellos operadores que tengan a 0 y a 1 como eigenvalores.

SECCION 6.1

1. (a) V (b) F (c) F (d) V (e) F (f) F
   (g) V (h) V
2. (a) Para $\lambda = 2$,

$$\left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right\}$$

es una base para E_λ; cualquier base para R^2 es una base para K_λ.

(b) Para $\lambda = -1$,

$$\left\{\begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}\right\}$$

es una base para E_λ y K_λ.

Para $\lambda = 2$,

$$\left\{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right\}$$

es una base para E_λ y

$$\left\{\begin{pmatrix}0\\1\\-1\end{pmatrix},\begin{pmatrix}1\\1\\1\end{pmatrix}\right\}$$

es una base para K_λ.

SECCION 6.2

1. (a) V (b) V (c) F (d) V (e) V (f) V
(g) F (h) V

2. $$\begin{pmatrix}2&1&0\\0&2&1\\0&0&2\end{pmatrix}\oplus\begin{pmatrix}2&1\\0&2\end{pmatrix}\oplus(2)\oplus\begin{pmatrix}4&1&0\\0&4&1\\0&0&4\end{pmatrix}\oplus(4)\oplus(-3)\oplus(-3)$$

3. (a) $-(t-2)^5(t-3)^2$
(b) Para $\lambda_1 = 2$ Para $\lambda_2 = 3$

(c) $\lambda_2 = 3$
(d) $p_1 = 3$ y $p_2 = 1$
(e) (i) $\text{rango}(\mathsf{U}_1) = 3$ y $\text{rango}(\mathsf{U}_2) = 0$
(ii) $\text{rango}(\mathsf{U}_1^2) = 1$ y $\text{rango}(\mathsf{U}_2^2) = 0$
(iii) $\text{nulidad}(\mathsf{U}_1) = 2$ y $\text{nulidad}(\mathsf{U}_2) = 2$
(iv) $\text{nulidad}(\mathsf{U}_1^2) = 4$ y $\text{nulidad}(\mathsf{U}_2^2) = 2$

4. (a)
$$J = \left(\begin{array}{c|cc}1&0&0\\\hline 0&2&1\\0&0&2\end{array}\right) \quad \text{y} \quad Q = \begin{pmatrix}1&1&1\\2&1&2\\1&-1&0\end{pmatrix}$$

(d)
$$J = \left(\begin{array}{cc|cc}0&1&0&0\\0&0&0&0\\\hline 0&0&2&0\\0&0&0&2\end{array}\right) \quad \text{y} \quad Q = \begin{pmatrix}1&0&1&-1\\1&-1&0&1\\1&-2&0&1\\1&0&1&0\end{pmatrix}$$

6. La forma canónica de Jordan es

$$\left(\begin{array}{ccc|c}1&1&0&0\\0&1&1&0\\0&0&1&0\\\hline 0&0&0&2\end{array}\right),$$

y una base canónica de Jordan es $\{2e^x, 2xe^x, x^2e^x, e^{2x}\}$.

SECCION 6.3

1. (a) V (b) F (c) F (d) V (e) F (f) V
   (g) F

2. (a) $\begin{pmatrix} 0 & 0 & 27 \\ 1 & 0 & -27 \\ 0 & 1 & 9 \end{pmatrix}$ (b) $\begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$

   (c) $\begin{pmatrix} \frac{1}{2}(-1+i\sqrt{3}) & 0 \\ 0 & \frac{1}{2}(-1-i\sqrt{3}) \end{pmatrix}$ (e) $\left(\begin{array}{cc|cc} 0 & -2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & -3 \\ 0 & 0 & 1 & 0 \end{array}\right)$

SECCION 7.1

1. (a) V (b) V (c) F (d) F (e) F (f) F
   (g) V (h) F (i) V

2. $(x, y) = 4 + i$, $||x|| = \sqrt{7}$, $||y|| = \sqrt{14}$, y $||x+y||^2 = 37$.

3. $(f, g) = 1$, $||f|| = \dfrac{\sqrt{3}}{3}$, $||g|| = \sqrt{\dfrac{e^2-1}{2}}$,

   y $||f+g|| = \sqrt{\dfrac{11+3e^2}{6}}$

SECCION 7.2

1. (a) F (b) V (c) V (d) F (e) V (f) V

2. (b) La base ortonormal es

   $$\left\{\frac{\sqrt{3}}{3}(1, 1, 1), \frac{\sqrt{6}}{6}(-2, 1, 1), \frac{\sqrt{2}}{2}(0, -1, 1)\right\}.$$

   Los coeficientes de Fourier son $2\sqrt{3}/3$, $-\sqrt{6}/6$, y $\sqrt{2}/2$.

   (c) La base ortonormal es $\{1, 2\sqrt{3}(x-\frac{1}{2}), 6\sqrt{5}(x^2-x+\frac{1}{6})\}$.

   Los coeficientes de Fourier son $3/2$, $\sqrt{3}/6$, y 0.

4. $S^\perp = L\{i, \frac{1}{2}(1-i), -1)\}$.

5. En el primer caso, $S^\perp$ es el plano que pasa por el origen que es perpendicular a x_0; en el segundo caso $S^\perp$ es la recta que pasa por el origen que es perpendicular al plano que contiene a x_1 y x_2.

SECCION 7.3

1. (a) V (b) F (c) F (d) V (e) F (f) V
(g) V

2. (a) $y = (1, -2, 4)$ (c) $y = 210x^2 - 204x + 33$

3. (a) $T^*(x) = (11, -12)$ (c) $T(f(x)) = 7x^2 + x + 12$

14. $T^*(x) = (x, z)y$

SECCION 7.5

1. (a) V (b) F (c) F (d) V (e) V (f) V
(g) F (h) V

2. (a) T es autoadjunto; la base ortonormal es $\left\{ \frac{1}{\sqrt{5}}(1, -2), \frac{1}{\sqrt{5}}(2, 1) \right\}$.

(b) T es normal pero no autoadjunto.
(c) T no es normal.

SECCION 7.6

1. (a) F (b) V (c) V (d) F (e) F

2. (a) $\sqrt{18}$ (c) 2

4. (a) $\|A\| \approx 84.74$, $\|A^{-1}\| \approx 17.01$, y $\text{cond}(A) \approx 1441$

(b) $\|\tilde{x} - A^{-1}b\| \leq \|A^{-1}\| \cdot \|A\tilde{x} - b\| \approx .17$ y

$$\frac{\|\tilde{x} - A^{-1}b\|}{\|A^{-1}b\|} \leq \text{cond}(A) \frac{\|b - A\tilde{x}\|}{\|b\|} \approx \frac{14.41}{\|b\|}$$

5. $0.001 \leq \dfrac{\|x - \tilde{x}\|}{\|x\|} \leq 10$

6. $R\begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} = \frac{9}{7}$, $\|B\| = 2$, y $\text{cond}(B) = 2$

SECCION 7.7

1. (a) V (b) F (c) F (d) V (e) F (f) V
(g) F (h) F (i) F

2. (a) $P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ y $D = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}$

(d)
$$P = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix} \quad \text{y} \quad D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

4. T_z es normal para toda $z \in C$; T_z es autoadjunto si y sólo si $z \in R$; T_z es unitario si y sólo si $|z| = 1$.

5. Solamente el par de matrices en el inciso (d) es unitariamente equivalente.

21. (c)
$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{\sqrt{6}}{6} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{\sqrt{6}}{6} \\ 0 & \frac{1}{\sqrt{3}} & \frac{\sqrt{6}}{3} \end{pmatrix} \quad \text{y} \quad R = \begin{pmatrix} \sqrt{2} & \sqrt{2} & \frac{2}{\sqrt{2}} \\ 0 & \sqrt{3} & \frac{\sqrt{3}}{3} \\ 0 & 0 & \frac{\sqrt{6}}{3} \end{pmatrix}$$

(e) $x_1 = 3,\ x_2 = -5,\ x_3 = 4$

22. (a) $x = \frac{1}{\sqrt{2}}x' + \frac{1}{\sqrt{2}}y'$ y $y = \frac{1}{\sqrt{2}}x' - \frac{1}{\sqrt{2}}y'$

La nueva forma cuadrática es $3(x')^2 - (y')^2$.

SECCION 7.8

1. (a) F (b) V (c) V (d) F (e) V (f) F
(g) F (h) V (i) V (j) F

3. $\left\{ t\begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix} : \ t \in R \right\}$

4. (b) $\left\{ t\begin{pmatrix} 1 \\ 0 \end{pmatrix} : \ t \in R \right\}$ si $\phi = 0$ y $\left\{ t\begin{pmatrix} \cos\phi - 1 \\ \operatorname{sen}\phi \end{pmatrix} : \ t \in R \right\}$ si $\phi \neq 0$

7. (c) Se tienen seis posibilidades:

(i) Cualquier recta que pase por el origen si $\phi = \psi = 0$.

(ii) $\left\{ t\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} : \ t \in R \right\}$ si $\phi = 0$ y $\psi = \pi$.

(iii) $\left\{ t\begin{pmatrix} \cos\psi + 1 \\ -\operatorname{sen}\psi \\ 0 \end{pmatrix} : \ t \in R \right\}$ si $\phi = \pi$ y $\psi \neq \pi$.

(iv) $\left\{ t\begin{pmatrix} 0 \\ \cos\phi - 1 \\ \operatorname{sen}\phi \end{pmatrix} : \quad t \in R \right\}$ si $\psi = \pi$ y $\phi \neq \pi$.

(v) $\left\{ t\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} : \quad t \in R \right\}$ si $\phi = \psi = \pi$.

(vi) $\left\{ t\begin{pmatrix} \operatorname{sen}\phi(\cos\psi + 1) \\ -\operatorname{sen}\phi \operatorname{sen}\psi \\ \operatorname{sen}\psi(\cos\phi + 1) \end{pmatrix} : \quad t \in R \right\}$ para cualquier otra posibilidad.

SECCION 7.9

1. (a) F (b) V (c) V (d) V (e) V (f) F
2. Para $W = L(\{(1, 2)\})$, $[T]_\beta = \begin{pmatrix} \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{4}{5} \end{pmatrix}$.
3. (ii) (a) $T_1(a, b) = \frac{1}{2}(a + b, a + b)$ y $T_2(a, b) = \frac{1}{2}(a - b, -a + b)$
   (d) $T_1(a, b, c) = \frac{1}{3}(2a - b - c, -a + 2b - c, -a - b + 2c)$ y $T_2(a, b, c) = \frac{1}{3}(a + b + c, a + b + c, a + b + c)$

SECCION 7.10

2. La parábola es $y = \dfrac{t^2}{3} - \dfrac{4t}{3} + 2$, y el error es 0.
3. $x = \frac{2}{7}, y = \frac{3}{7}, z = \frac{1}{7}$

SECCION 7.11

1. (a) F (b) F (c) V (d) F (e) V (f) F
   (g) F (h) F (i) V (j) F
4. (a) Sí. (b) No. (c) No. (d) Sí. (e) Sí.
   (f) No.
5. (a) $\begin{pmatrix} 0 & 2 & -2 \\ 2 & 0 & -2 \\ 1 & 1 & 0 \end{pmatrix}$ (b) $\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$ (c) $\begin{pmatrix} 0 & 0 & 0 & 0 \\ -1 & 0 & -4 & 0 \\ 0 & 0 & 0 & 0 \\ -2 & 0 & -8 & 0 \end{pmatrix}$

22. (a) $$\left\{\begin{pmatrix} \frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix}\right\}$$

(b) La misma que (a).

(c) $$\left\{\begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix}\right\}$$

23. Igual que 22(c).

33. (a) $Q = \begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix}$ y $D = \begin{pmatrix} 1 & 0 \\ 0 & -7 \end{pmatrix}$

(b) $Q = \begin{pmatrix} 1 & -\frac{1}{2} \\ 1 & \frac{1}{2} \end{pmatrix}$ y $D = \begin{pmatrix} 2 & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}$

(c) $$Q = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & -0.25 \\ 1 & 0 & 2 \end{pmatrix} \quad \text{y} \quad D = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6.75 \end{pmatrix}$$

Lista de símbolos usados frecuentemente

Indice alfabético